DOCUMENT RESUME

ED 108 259 CS 202 148

AUTHOR Knudson, Richard L., Ed.
TITLE Emphasis: Use and Misuse of Standardized Testing.
INSTITUTION New York State English Council.
PUB DATE 75
NOTE 114p.; Some pages have small type which may result in poor reproducibility
AVAILABLE FROM Alan Nelson, Executive Secretary, Union College, Humanities Center, Schenectady, New York 12308 ($2.00)
JOURNAL CIT The English Record; v26 n2 Entire Issue Spring 1975

EDRS PRICE MF-$0.76 HC-$5.70 PLUS POSTAGE
DESCRIPTORS Elementary Secondary Education; English Curriculum; *English Instruction; *Evaluation Methods; Language Arts; *Language Development; *Measurement Instruments; Objective Tests; Reading Instruction; *Standardized Tests

ABSTRACT

Because of the active interest of English teachers in standardized tests and the increased use of such tests to measure student performance, the emphasis section of this issue of "The English Record" contains articles addressing this subject from several perspectives. A report of a survey on standardized testing in New York State sets the scene, demonstrating that English teachers feel a need for test information, and pointing up the necessity for educating the general public in the proper use of test scores. The articles that follow define terms used in testing and discuss the issues. Questions are raised about some of the tests in wide use throughout the state of New York, and several articles offer alternatives to traditional ways of evaluating reading and writing. The emphasis section concludes with a recent sources bibliography on measuring growth in English. The general articles contained in this issue discuss the heterogeneous grouping of high school students, the various uses of the cloze technique in the teaching of English as a second language, Robert Frost's "The Road Not Taken," the development of a competency-based English curriculum in kindergarten through grade twelve, and the advantages of using improvisation in the English class. (RB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED108259

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

the english record

Official Publication of the New York State English Council
Copyright New York State English Council 1974
Second Class Postage Paid at Schenectady, N. Y.

Editor: Richard L. Knudson

Associate Editors: Elizabeth Drake, David Stineback, Marjorie Rogers, Charles Cooper

---

VOL. XXVI **SPRING, 1975** **No. 2**

TABLE OF CONTENTS

EDITOR'S PAGE ---------- 2

EMPHASIS: USE AND MISUSE OF STANDARDIZED TESTING

Introduction ---------- *Ouida H. Clapp* 3
Results of the NYSEC Survey on Testing ---------- *Robert Infantino* 4
Evaluating and Liberating Teaching and Teachers ---------- *Morris Finder* 11
Criterion Tests and Standardized Tests ---------- *Morris Finder* 14
Limitations of Standardized Tests as they Apply to Measuring Growth in English Curriculum Areas ---------- *La Ruth Hackney Gray* 17
The Regents Comprehensive Examinations in English: Some New Questions and Some Tentative Answers ---------- *Lee Odell* 27
A Proposal for a Test in English For Students in Grades 11-12 ---------- *Robert Carruthers* 38
The Cooperative Development of Growth Measures for the Ability to Read Fiction ---------- *Thomas F. Callaghan* 42
Are Six Weeks Gone Already? Or An Alternative to Measuring Growth in Writing Ability ---------- *Kenneth Watson* 49
Grades and Growth in the Writing Program ---------- *Charles R. Cooper* 56
Recent Sources for Measuring Growth in Writing and Reading Fiction ---------- *Thomas G. Mac Lennan* 67

GENERAL ARTICLES

R + NR = ? ---------- *Patricia E. Moore* 78
John T. Chamard
Putting Cloze into the Classroom ---------- *John F. Haskell* 83
Reading Frost: "The Road Not Taken" ---------- *Roberts W. French* 91
A Development Cycle for a Competency Based English Curriculum, Grades K-12 ---------- *Lester S. Golub* 94
Improvisation and the Teaching of English ---------- *Patrick L. Courts* 101
Literary Research ---------- *R. D. Lakin* 111
Media Review ---------- 112

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
New York State
English Council
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

CS 202 148

Member of the NCTE Information Agreement

THE ENGLISH RECORD is published quarterly (fall, winter, spring, and summer) by the New York State English Council. Editorial Office: Richard L. Knudson, State University College, Oneonta, New York 13820. General articles of interest to English teachers on all levels of instruction are invited. Manuscripts should be sent with return postage; whenever possible, documentation should be included in the text. Letters to the editor as well as book reviews are welcome. Membership (including a year's subscription to The English Record) $10.00 annually. Individual issues: $2.00 each. Correspondence, subscriptions and advertising should be addressed to Alan Nelson, Executive Secretary, Union College Humanities Center, Schenectady, New York 12308.

— Editor's Page —

At this time of the year teachers become anxious about evaluating their work. Many of us look at the final testing period as a means of evaluating what students have done. The answer to the question, "Can we find out what's in someone's head?" remains, "We can't."

The *Emphasis* section this month has been assembled by *The Record's* Associate Editor, Charles Cooper, and deals with the problems of evaluation and testing. We are particularly proud of this section because it represents a wide range of thought on the matter and should help to clarify (or start) thinking on the subject.

I am very pleased about the quality and quantity of manuscripts I receive. Each goes through several careful readings before being accepted or rejected. If the journal were a monthly instead of a quarterly, then we could accept more of them. Since this is not so, many excellent manuscripts are sent back to the writers. Manuscripts chosen for publication are not only of high quality but also are timely. Please do keep submitting those manuscripts. I am particularly interested in seeing more articles from elementary school teachers. Time, of course, is a problem, but summer is coming.

Lastly, one reads a great deal these days about going "back to basics." What do you think? Is there a movement growing in New York State for this? I would like to see three or four articles in the fall issue on this subject. If it intrigues you, then why not write about it?

INTRODUCTION

High on the agenda of items of active interest to English teachers is the matter of standardized tests, their proliferation and their mushrooming role in English education. Many teachers are asking whether the accelerating use of such tests is accompanied by a general understanding of the strengths and limitations of the tests and by a clear perception of their proper role in English instruction and evaluation.

Alert to the questions English teachers are raising, the New York State English Council appointed a Committee on the Use and Misuse of Standardized Tests. The committee's work has included sponsoring a statewide survey to determine what English teachers know and how they feel about standardized tests as well as how teachers use them; studying ways of disseminating test information to members of NYSEC; and cooperating with the National Council of Teachers of English in the development of policy concerning standardized tests.

In this issue of *The English Record*, the committee presents a group of articles that address the subject from several perspectives. A report of the survey sponsored by the committee sets the scene, demonstrating that English teachers feel a need for test information and pointing up the necessity for educating the general public regarding the proper interpretation of test scores. The articles that follow define terms used in testing and discuss the issues. Questions are raised about some of the tests in wide use throughout the state. Several articles offer fresh, provocative alternatives to traditional ways of evaluating reading and writing. A "recent sources" bibliography closes the series.

The members of NYSEC's standardized testing committee are:

Charles R. Cooper
State University of New York at Buffalo

Morris Finder
State University of New York at Albany

La Ruth Gray
Isaac E. Young Junior High School
New Rochelle, New York

Robert I. Infantino
Committee Researcher
State University of New York at Buffalo

It has been a personal pleasure to be associated with the committee in this important effort, and I wish to express my sincere appreciation to its members for the time, effort, and valuable insights they have contributed so generously to our common task.

Ouida H. Clapp
Buffalo Public Schools

Results of the NYSEC Survey on Testing

by Robert Infantino

In May of 1974 a survey was conducted among nearly 1000 randomly selected English teachers across New York State to ascertain current practices in standardized testing and how English teachers felt about these practices. The New York State English Council (NYSEC) Committee on the Uses and Abuses of Standardized Testing sponsored the survey, and responses were collected from 509 English teachers. This article will present some of the findings of that survey.

Using the computer resources of the Bureau of Educational Data Systems (BEDS) of the New York State Education Department, a random sample of 985 English teachers was compiled. This sample included teachers from all public school systems in the State, including the New York City teachers. Of the 509 teachers who responded (51.6%), there were 497 valid, useable surveys, a return of 50.5%. These included 152 who identified themselves as teachers in city schools, 243 in suburban schools, and 100 in rural schools. The figures do not always equal 497 since some respondents omitted some items.

There were 295 senior high school teachers (grades 9-12) and 174 junior high or middle school teachers (grades 6-9). Twenty-eight (28) taught in other situations which included teachers of grades 7-12 or some teachers in special language arts programs in elementary schools. A wide range of teaching experience was represented by the respondents, with 69 being in their first or second year of teaching and with 95 having more than fifteen years as teachers.

The survey itself asked teachers to respond to 37 items concerned with their knowledge of current testing practices in their own school systems and how these practices affected their English classrooms. Testing was defined as either standardized, i.e., scored on the basis of a set of norms for a large population (like the SAT), or externally prepared, i.e. prepared by someone other than the teacher of the course (like the Regents Comprehensive English Examination).

The second part of the survey included 27 attitudinal items on which teachers were asked to indicate their feelings about testing and some of its ramifications such as accountability, gain in pupil performance as a measure of teacher effectiveness, and cultural bias in standardized tests. Complete results of this survey cannot be printed here but will be available through University Microfilms and possibly through NYSEC. What follows is a limited reporting of some of the data collected, with comments about the significance of some of the responses.

There were 427 teachers who said that the scores pupils receive on standardized tests were available to the teachers at nearly any time. Only 2 said that scores were not available to teachers at all, while others indicated that an administrator or counselor must give permission to see the scores. Yet only 294 of the teachers stated that they feel competent enough to interpret test scores themselves and only 173 have ever attended meetings or workshops intended to acquaint teachers with the use and interpretation of standardized tests. These statistics indicate a need for some in-service work in school districts to assist their teachers in more fully understanding and properly using the test results which appear to be readily available.

Teachers were next asked to check all of the tests which their school systems administered to pupils in grades 7-12 in four categories: Intelligence, Standardized Achievement, Reading, and Standardized Aptitude Tests. The tests cited most frequently in each of the categories were these:

Intelligence—Stanford-Binet (146), and the California Test of Mental Maturity (106);

Standardized Achievement—College Board Achievement Tests (232), Iowa Tests of Education Development (194), New York State Pupil Evaluation Program—PEP (167), and the Stanford Achievement Test (153);

Reading—SRA Achievement Series: Reading (153), Metropolitan Achievement Tests: Reading (152), and Gates Mac Ginitie Reading Tests (110). (82 respondents said that other reading tests besides the six listed in the survey were administered in their school systems);

Standardized Aptitude—Scholastic Aptitude Tests—SAT (318), Preliminary Scholastic Aptitude Tests—PSAT (309), Regents Scholarship Qualifying Tests (290), and Differential Aptitude Tests—DAT (129).

Several inferences can be made from the data above as well as from some of the other responses in these categories. For instance, other than the SAT and the PSAT, no test was cited by more than 50% of the teachers. This indicates the great multiplicity of testing programs throughout the state and further points out the need for explanatory sessions particularly for new teachers—or for experienced personnel when new testing programs are introduced.

The New York State Pupil Evaluation Program (PEP) has been in existence for the past four years in all school systems in the State. Even though it is mandatory at the ninth grade level, only 167 respondents recognized it as being administered in their school systems in grades 7-12. While newspapers, particularly in large cities, frequently cite the results of this test as an indicator of the success or failure of the schools, only one-third of the teachers listed this test on the survey.

In addition to the above results, there were many teachers in each category who indicated that they did not know whether their school system administered tests in grades 7-12: 76 did not know about intelligence testing, 49 about achievement testing, 47 about tests in reading, and 58 about tests of aptitude. Some kind of information program about school testing programs is surely indicated.

Three items on the survey pertained to helping pupils prepare for such standardized tests as the SAT or the PEP. 89 respondents stated that their school systems required English teachers to spend class time in special preparation for such tests. In a later item, 183 teachers agreed that it is a legitimate use of instructional time in English class to help pupils prepare for these tests, while 220 teachers disagreed with that practice. However, there were 281 who indicated that they felt coaching pupils could significantly raise the pupils' scores on the SAT. Only 82 disagreed with this item while 127 were undecided, even though test makers and educational researchers and writers note that, overall, coaching will not alter pupils' scores very much except to make them more "test-wise," i.e. ready for the mechanics of the particular test.

The practice of tracking or grouping pupils in English on some pre-determined basis was surveyed. The question was, "How many tracks are there in your school's English program?" Three tracks was the response of 171 people; four tracks had 96 responses; two tracks were indicated by 92 teachers; five or more tracks were listed by 44 respondents; one track was cited by 8 people. There were 75 teachers who stated that their school's English program had no separate tracking system.

A total of 386 teachers indicated that pupils were grouped or tracked on the basis of both standardized tests and other factors such as grades or teacher recommendations. 42 others stated that pupils were grouped solely on the basis of standardized tests of IQ and/or reading. Once again the proliferation of practices shows a real need to study the situation to be sure that whatever is being done in the school system has a purpose, and that the purpose for grouping pupils is rational and educationally sound, and is done without prejudice toward any group or individual.

There were several items pertaining to the New York State Regents Comprehensive English Examination. Nearly 60% of the respondents (296) indicated that their school system has a separate curriculum or track for those pupils preparing to take the Regents. In addition, 95 teachers cited the existence of a separate elective course which pupils take to prepare them for the Regents. When asked about the proportion of pupils in their school system who take the Regents exam, 95 teachers said that less than 50% of the pupils take the exam, while 281 teachers said that more than half the pupils do take the exam at some point in their high school years. There were 153 teachers who did not know the proportion of pupils taking the exam in their school system.

In response to an attitude question about the Regents examination, 257 teachers (53%) indicated that they felt that the Regents exam has a restrictive influence on the curriculum for the students who will eventually take the exam. Only 175 (35%) did not feel this restrictive influence while 53 (11%) were undecided. In the continuing controversy over the value of administering the Regents exams in New York State, these figures seem to show that English teachers would rather not have the exams since they tend to restrict the school's curriculum. More is written about the Regents and the curriculum later in this article and elsewhere in this issue.

Regents examination scores are used in a variety of ways in the various school systems. 286 teachers mention that Regents scores are included in the pupil's end-of-the-year final average in English. Another 134 teachers indicate that Regents scores are used solely for the purpose of the pupil's obtaining a Regents diploma and are not counted toward any average or end of course evaluation. The Bureau of English Education of the New York State Education Department indicates that the choice of how to make use of the scores is left to the school system, although any pupil may attempt the examination even though not specially prepared or grouped in a Regents track.

The most common examination practice in English classrooms in New York is the practice of the classroom teacher(s) of that course or grade level preparing the exam for the pupils in the school. There were 299 teachers (61%) who indicated this on the survey. Other responses on end-of-course exams included 72 which were prepared by a school committee, 8 by department chairmen, 15 by a district committee, and 7 by the central office staff. There were 83 teachers who felt that the question was not applicable to their school systems, probably because they prepared their own exams or because no final exams were given. Later in the survey, teachers expressed an overwhelming preference for two types of end of course exams—self prepared (57%) or cooperatively constructed within one school (27%). Only 35 teachers (7%) indicated a preference for the Regents or some other state-wide exam.

When asked about the *quality* of end of course exams not solely prepared by the teacher of the course, 80 teachers indicated satisfaction, 133 cited the need for improvement, and 54 indicated that the quality was unsatisfactory. This item was followed by one which asked if teachers felt that the content of their English course was restricted by externally prepared exams like the Regents or a district-wide exam. The teachers were nearly evenly split on this item, 245 saying yes they felt restricted in content, with 239 saying they did not feel restricted. The split was a little different when the teachers were asked if they felt that externally prepared exams restrict the way they conduct instruction in English class—206 said that they felt a restriction in methodology while 270 said that they felt no such restriction.

To find out what teachers might add to their English courses if they felt no restrictions because of standardized tests or externally prepared exams, two questions were included in the survey. One dealt with the *content* of the courses, with 164 teachers citing values clarification as the content they would add. This was followed closely by modern fiction (158), film study (156), and reading instruction (152). Feminist literature (57) and formal grammar instruction (71) were the content additions least favored. There were 141 teachers who would make no changes in the content of their courses.

The second question dealt with the *activities* which teachers would emphasize more strongly in their English courses if there were no mandated tests. Creative personal writing was the activity cited by the most teachers (222). Other activities which received emphasis were individual student projects (201), dramatic activities (182), and small group discussions (181). There were 134 teachers who said they would not change the emphasis of the activities in their English courses.

The question of teacher accountability for pupil learning is one which has been raised in recent years. There has been much writing both for and against the accountability movement in education, some saying that accountability will produce nothing but test taking pupil-robots, while others argue that accountability is vital if public education wants to continue to receive or to increase its share of the public tax dollars.

The survey asked the question, "For which of the following should teachers be held accountable?" (check all you feel applicable). The 497 respondees answered as follows:

467 their subject matter knowledge (94%)
426 knowledge of the factors that influence learning (86%)
398 the use of professionally sound instructional procedures (80%)
79 measurable growth in pupil learning based on standardized test scores (16%)
280 measurable growth in pupil learning based on accomplishment of the teacher's instructional objectives (56%)
306 specification of desired instructional objectives for their classes (62%)
429 satisfactory demonstration of classroom teaching skill (86%).

It might be said on the basis of the above figures that New York State's English teachers are willing to be held accountable for those factors over which they have some control—knowing their subject matter, demonstrating their teaching skill, specifying objectives for their classes. They are also willing to be accountable for using sound instructional procedures and knowing (and applying) the factors that influence learning in their classrooms. On the other hand, teachers reject being held accountable for pupil learning based on standardized test scores, probably because they know or intuitively feel that these test scores are based on more factors than just the teacher's own classroom abilities or even on the pupils' knowledge of facts presented in English classrooms.

There is some uncertainty shown in the responses to the item asking whether the teachers would be accountable for growth in pupils based on the teacher's instructional objectives. Later in the survey the teachers were asked if they felt it was possible to specify instructional objectives in measurable terms for all areas of English instruction. Only 102 teachers (21%) agreed that it was possible, while 336 teachers (68%) disagreed with the possibility. It appears that the willingness to write objectives is somewhat tempered by the belief that some of English classroom instruction cannot be stated in objective, measurable terms. This is a dilemma which merits more investigation and discussion as the accountability/objectives questions become issues in more and more school systems.

Several more items on the survey dealt with teacher attitudes toward accountability in the profession. There were 331 teachers (67%) who felt that they could support an accountability program if its goal was to improve instruction in English. They also felt (398—80%) that success in school achievement should be measured by results achieved in the actual situation of the particular school and the particular set of pupils rather than by a standardized test.

52% of the teachers (261) stated that they would continue in the school system if their local teacher association agreed to an accountability program, even if the teacher himself did not like the program. Significantly, there were 169 respondees (34%) who were undecided about this item. When asked if they would attempt to find a position in another school system if their system imposed an accountability program on its teachers, only 51 (12%) said that they would. 227 teachers (46%) said they would not try to find another position, while 194 (39%) were undecided. The small number who said that they would try to move should encourage school systems on the issue of accountability, particularly in view of the shortage of teaching positions in English. There would undoubtedly be many other factors and pressures brought to bear on a system which imposed an accountability program on its staff. Yet if these figures reflect the feelings of a broader range of teachers, who knows what is possible?

Those who responded to the survey were consistent in their feelings against using standardized instruments to measure pupil achievement in English. A total of 399 teachers (80%) stated that they felt standardized achievement tests are not adequate measures of pupil growth in an accountability program. Another 288 (57%) disagreed with the statement that gain in pupil achievement based on a pre-test, instruction, and a post-test is a satisfactory method of evaluating a teacher in an accountability program. In fact, 240 teachers (48%) felt that there is no satisfactory method for evaluating a teacher on the basis of how well or poorly his pupils perform. Only 139 (28%) felt there was a way of accomplishing this evaluation, and 109 (22%) were uncertain.

There were a few other issues and attitudes covered in the survey, but those reported here seemed to be central to the interests of the profession. Standardized testing has grown into a multi-million dollar business in this country. English teachers and all other educators, school board members, and parents need to pay close attention to the uses of tests in their schools. Teachers need to be better informed about the testing programs their particular system administers, and they need to speak out when they feel that the pupils or the teachers are being somehow abused by the givers or interpreters of any test or series of tests.

This survey reflects some strong, negative attitudes in the English teachers of New York State, particularly toward the Regents Comprehensive English Examination and toward the use of standardized testing to measure teacher effectiveness with pupils. It reflects a great lack of information on the uses of tests in many school systems. There is also an indication of the great variety of testing programs and organizational patterns in English classes as a result of the testing programs.

In spite of all of the confusion and misapprehension, however, 401 of the respondents, a huge 80%, agreed with the statement, "Teaching English in public schools is a good career choice for me." Only 18 (4%) said that it was not, while 66 (13%) remained undecided. This speaks very highly of the state of the secondary school English teachers of New York who apparently enjoy their teaching careers even while wrestling with all of the complex problems and issues discussed in this survey.

SUNY, Buffalo

NEW YORK UNIVERSITY / SCHOOL OF EDUCATION

GRADUATE STUDY ABROAD SUMMER SESSION 1975

Earn 12 graduate credits in England /
Dates: June 30-August 22, 1975 / **Tuition**: $720

For Teachers of English: at the University of York, England. For the initial summer, we are seeking English and language arts teachers interested in learning new approaches and in developing new curricula for their classrooms. Stressing the personal growth model of English, our staff will be led by Geoffrey Summerfield, and the program will include observation and participation in British schools.

For applications and further information, write
Office of Off-Campus Programs (NCTE)
School of Education
New York University
23 Press Annex, Washington Square
New York, N.Y. 10003

Please include NCTE in reply.

Evaluating and Liberating Teaching and Teachers

by Morris Finder

This issue of *The English Record* is devoted to testing, which is an aspect of evaluation. An understanding of evaluation is therefore prerequisite to an understanding of testing. Both testing and evaluation are means to the improvement of teaching and to the intellectual and personal freedom of teachers. To justify these assertions, this article describes evaluation and its procedures and uses; it explains also why and how it offers teachers increased freedom.

Evaluation of teaching or of anything else is a process of determining the extent to which the purpose has been or is being achieved. If, for example, the purpose of a hardware store is profit, then we evaluate by assessing its profit or loss. Similarly, if the purpose of teaching is the learning of students, we evaluate by finding out the extent to which the teaching has achieved the desired learning. From this purpose, the procedures of evaluation may be inferred. The following list, then, specifies the procedures.

1. Obtain a statement of the program's instructional objectives. If one is not available, formulate it.
2. Select or devise exercises, problems, or other tasks that give the student the opportunity to exhibit the achievements specified by the objectives.
3. Administer the set of evaluative tasks.
4. Obtain a record of the student's performance.
5. Determine from the record the extent to which the performance exhibited conforms to the performance desired.

To illustrate the uses of these procedures, we take as an example the objective, "Identify devices of persuasion in advertisements." With the objective stated, we then go on to select situations or problems that give the student the opportunity to exhibit the extent to which he has acquired that achievement. It is easy to infer that the evaluative problems would be a set of advertisements with some such instruction as this: "For each of these ads, identify the devices of persuasion incorporated into it." We next administer this set of problems. The students record their own performances as a consequence of taking the written test implied. The teacher analyzes the results to determine strengths and weaknesses of the performances and, thereby, of the teaching. For example, the results may show weaknesses in ability to perceive this or that device of persuasion or weaknesses in these related skills, and so on. From such information, the teacher makes changes in the program, teaches it again, evaluates, revises, and so on without end. Planning and teaching, then, is a continuous, cyclical process.

Mastery learning exemplifies evaluation used to improve instruction. Tests that are written from explicit instructional objectives are sometimes called mastery tests and sometimes criterion tests. Another article in this issue describes criterion tests and compares them with tests called standardized or norm-related.

Benjamin S. Bloom and his associates at the University of Chicago have been investigating the possibilities and problems of mastery learning. The assumption behind it is that some 90% of students can learn what is taught provided the teacher identifies explicit instructional objectives, uses appropriate means to obtain the learning that the objectives specify, evaluates often, and adjusts the instruction from the information yielded by the procedures of evaluation.

Bloom gives good reasons for rejecting the familiar practice of grading on a curve. The bell-shaped curve, he reminds us, is the model for random activity and chance. Because teaching is a purposeful activity, there is reason to suppose that teaching is unsuccessful if achievement corresponds to the model of random activity.[1]

There can be no more important use of evaluation than the improvement of instruction. A unit of instruction is likely to be improved if evaluation occurs at least three times: at the beginning to assess the learners' strengths and weaknesses, toward the middle to measure progress, and at the end to determine the success of the entire program.

Although we have used a paper and pencil test to exemplify an evaluative instrument, evaluation cannot be limited to such means. If, for example, an objective concerns public address, then a valid evaluation must put the student to work addressing an audience. The record of evaluation, then, would be provided by those evaluating, probably assisted by criteria that define satisfactory performance.

For another example, if the objective concerns effective group discussion, the corresponding evaluation implies a group situation with a record to be made by those observing the situation. Only in paper and pencil tests do students record their own performances.

Not only is evaluation grounds for improved teaching, but its results are informative to the administration, the community, and, sometimes, to the profession.

Because a proper evaluation is the means for determining the success of a program, it follows that this success is in principle independent of opinions about it. If, then, a teacher has planned and executed a program that sound evaluation shows to be successful, that teacher is thereby successful, regardless of positive or negative opinion. From this we infer the ultimate principle by which to establish both the freedom and responsibility of the teacher: He is successful to the extent that he achieves the ends to which his teaching is a means.

This article, then, has explained the evaluation of teaching and has suggested its uses in improving instruction, in informing the administration, the community, the profession, and in promoting the freedom of the teacher.

NOTE

[1] Benjamin S. Bloom, "Learning for Mastery," **Handbook on Formative and Summative Evaluation of Student Learning**, ed. Bloom, Hastings, and Madaus (New York: McGraw-Hill Book Company, 1971), pp. 11-15.

SUNY, Albany

Syracuse University
syracuse summer 75

WRITING FOR CHILDREN
Advanced Study in Children's Literature
JULY 14 to 25, 1975

For teachers, librarians, parents, writers, and prospective writers and/or illustrators of books for children, an opportunity to share ideas with authors, illustrators, editors, librarians, reviewers. Consultants will include Natalie Babbitt, author of

Knee-Knock Rise

Advanced registration is desirable. Suggestions, booklists, requirements will be mailed in advance to registrants on request.

Audit or Graduate Credit (EDU 760 Workshop in Education)

Write to: Dr. Margaret Early
252 Huntington Hall
150 Marshall Street
Syracuse, New York 13210

Criterion Tests and Standardized Tests[1]

by
Morris Finder

To understand current controversies about testing, one must know

1. what normed or standardized tests are,
2. what criterion tests are, and
3. what the differences between these are.

One's knowledge of these distinctions is prerequisite to the intelligent conduct and judgment of schools and teaching.

Let us consider first the normed or standardized test. A brief historical note can help. In 1914 there were about 139,000 soldiers in the United States Army. With the onset of World War I, the army grew rapidly to about two million. Therefore the army was faced with an immense task of sorting. Who should go to officer candidate school? Who should go to cooks' and bakers' school? Who should be selected for this kind of job? And who for that? These decisions had to be made promptly and so pyschologists were put to work on the problem. One result was the *Army Alpha Test*. When this test was administered to an unselected group, the army was told who was high, middling, and low on this particular instrument that measures some notion of mental alertness.

When the war had ended, some of the psychologists who developed these tests took jobs in colleges and universities. They taught the techniques of test construction they had developed in the army and adapted these to civilian educational uses. A result is that today such normed or standardized tests are widely believed to define educational testing and are often called, and believed to be, achievement tests. The following discussion shows, however, that both of these popular notions are dubious and a cause of some serious misunderstandings.

We begin by explaining why standardized testing cannot be equated with educational testing generally. If you had to write a standardized test, you'd have to devise items that half the typical takers of the test could not do. If the typical student doesn't fail half the items, then the test isn't functioning as a standardized test must.

It is easy to see, then, that the basis for writing a standardized test is not what students know or can do. Rather the purpose is to determine how one student compares with another on a scale from high to low.

But schools and teaching are intended to have students know things and be able to do things. It would follow that for practical teaching, the right kind of test is one that gives the student the opportunity to display the extent to which he can do the things he has been taught to do, understand, appreciate, and so on. Such instruments of evaluation get at desired achievement. Despite this, we find that school systems typically measure and report educational growth *not* with tests designed to reflect specifically what students have been taught but rather with tests intended to discriminate among students.

We use the term "criterion test" or "criterion-related" or "criterion-referenced" to name tests that measure directly and specifically the intentions of teaching. In such tests, norming or standardizing is not involved. And the test discriminates only on such matters as whether the student possesses the achievements desired. To write the test is to elicit the skills taught. For a criterion test, such skills and other behaviors are the criteria of test construction. That explains such terms as "criterion-related" and "criterion-referenced," and why it is reasonable to call such instruments achievement tests.

It is easy to see, then, the dubiety of typical reports of school systems. A school system may report, for example, that the achievement in reading of its students is good, bad, or indifferent. Typically the data are based on results of this or that standardized test. For reasons given, the validity of these reports, including references to grade-level equivalents, are dubious and misleading; it is fair to conclude that such misleading reports exemplify a misuse of standardized tests.

The "Coleman report," and the "Jencks report" exemplify a second and related misuse. In the winter, 1974 issue of *The Public Interest*, Ralph W. Tyler notes that the Coleman report on *Equality of Educational Opportunity* and Christopher Jencks' book, *Inequality*, claim that schools are relatively ineffective in teaching the disadvantaged. Tyler points out, however, that

> Both the Coleman and the Jencks studies examined differences in scores on standard tests among different groups of children. They did not ask what different groups of children had learned but rather what measured variables (i.e. socio-economic status) were related to differences in scores. The standard tests used were norm-referenced tests. In building these tests, questions that most children could answer correctly were eliminated, but questions which only about half the children could answer correctly were retained. This was done in order to spread the scores as widely as possible so that children could be arranged on a scale from highest to lowest. The purpose of norm-referenced tests is to sort students, not to assess what they have learned. It happens that many of the items that are effective in sharply sorting students are those that are not emphasized in a majority of schools.[2]

Tyler goes on to note that by age 13, 80 per cent of American children can read and comprehend a typical newspaper paragraph. An exercise such as this is included in the National Assessment of Educational Progress. The purpose of this National Assessment is to report what proportion of children has acquired this and other useful skills that schools do teach. Such an exercise is not included in standardized tests for 13-year-old children "because it does not sharply separate the very skillful reader from others."[3] Coleman and Jencks were using these standardized tests because they show the largest differences among groups. They found that family background was more related to these differences than the effects of the school were. But neither the test data nor the method of analysis of variance that they used could answer the question of what most children had learned in school.

It follows that if we want to know how good schools and teaching are we must use tests that measure what schools and teachers teach. The tests must be so written that if the student has learned, the test will show it. Teachers can write such tests and put them to constructive uses. The principle for writing the test is simple and obvious: Begin with statements of what the student is expected to know or do. Then write items that give him the opportunity to exhibit these desired achievements. If, for example he is to be able to state the literal sense of what is going on in a lyric poem, then present him with a lyric poem that presumably he hasn't seen before but is within the range of his experience and ask him to explain what, literally, is going on in it.

Standardized tests, however, do have their uses. If there is a need for sorting students on a range from low to high, then a sorting kind of test is appropriate. College admissions tests exemplify one of these uses and the Graduate Records Examination exemplifies another.

The purpose of teaching, however, is not to sort students but rather to help them achieve. Because desired achievements are the criteria for writing criterion tests, these are proper for typical classroom uses.

To summarize: Standardized tests are intended to sort people out, not to elicit learning sought by particular teachers and schools. It is misleading, therefore, to use the results of standardized testing alone to represent the achievements of educational programs. It is also misleading to use standardized tests to compare one school or school system with another: such comparisons are not necessarily based on the skills or other kinds of behavior that any or all of the schools involved have taught. A program of instruction is properly assessed by criterion tests because these are based on the skills taught and so written that the students to be tested are given an opportunity to demonstrate the extent to which they have acquired the desired learning. The results of such tests provide the primary acceptable bases for improving our schools and our teaching.

NOTES

[1] Revised version of a talk presented to a session of the Twenty-Fourth Annual Conference of the New York State English Council, Binghamton, New York, May 4, 1974.

Ralph W. Tyler, "The Federal Role in Education," **The Public Interest,** No. 34, Winter 1974, 170.

Ibid. 170-71.

SUNY, Albany

Limitations of Standardized Tests As They Apply to Measuring Growth in English Curriculum Areas

La Ruth Hackney Gray

Prior to ESEA funding in the late fifties and early sixties, standardized testing had basically served the interest of the practitioners in the school setting—teachers, administrators, college admissions officers, and research and development staff.

The late fifties and early sixties brought to the attention of national leaders during Congressional debate on several massive aid-to-education bills, the need for dependable information in evaluating cost benefits of cognitive growth and other related factors. The route that most research and development organizations, hired by programs for evaluation, took was monitoring through pre- and post-standardized achievement tests.

Once legislative debates in state and national capitols began, public attention, largely through the interest of the media, began to focus on "those tests."

"Those tests," standardized tests, are tests composed of empirically selected materials. They have definite directions for use, adequately determined norms, and data on reliability and validity.[1]

The norming of all standardized tests (achievement and aptitude) yields the average performance in a test of children of a certain age or in a certain grade. Norm referenced tests are constructed specifically to facilitate making comparisons among students.

CONTROVERSY OVER STANDARDIZED TESTING

As Leona Tyler points out, by 1970 public attitude had strongly developed with doubts about the "social utility of the whole testing enterprise." The testing industry has been subjected recently to a national wave of disenchantment, skepticism, and hostility as evidenced in numerous law suits, court rulings, and in the position taken by the Association of Black Psychologists and the American Personnel and Guidance Association.

The furor stems basically from two major irritants. The first: Standardized testing has been utilized primarily by our institutions to "wall" some students in and to "wall" some out; to provide schooling only for some children and education for others; to predetermine life styles and careers for select groups of the population; to wrongfully classify and label segments of our school populations; and finally to present arguments (at highly placed sources) for heredity and genetic endowment as the predominant determinant in explaining the consistent differences in obtained means between test results of majority and minority populations.

Evidence offered by Bloom et al. (1971); Goldberg, Passow and Justman (1966); Goodlad, (1966); Eash, 1961; Borg, 1964; Schwebel, 1968 negates the use of standardized tests to predict and select talent.

The second irritant is inherent in the design of standardized tests. Obviously, the scores do not represent what a student has learned but where he stands in relation to other students. Traditionally among test makers, the items selected are those found in the tryouts to differentiate most sharply among students, items that approximately half the students answer correctly.

In the language of reporting, grade equivalent misinterpretations are problematic and misused. Grade equivalent means that on the specific test the number of items that the student answered correctly is equal to the *median* of the number of items answered correctly by all students in the norming group. E.g. It is possible for a student in the 8th grade to gain 8.5 in comprehension on the Iowa Test of Basic Skills and not have mastered the inferential skills. Depending on the spread in raw score points, ranges can amount to three months growth on one part of the scale and three years on another part. Units are not equal throughout the scale.

The design of standardized tests by their very nature must penalize to some extent linguistic, experiential and cultural differences. The evidence is strong that all tests, published and unpublished, are biased to some degree.

Factors: *Congruence*—The response must match the preconceptions of the test maker. There is no allowance for divergent thinking. *Procedures used in item tyrouts*—These procedures create further problems, as explained by Green (1971):

> There is a biased effect when data from the population, or a sample of it, are used to improve the effectiveness of the test by selecting, rearranging, and rewriting items. This procedure is essential to producing an effective achievement test, but the improvement derived from it is not uniformly beneficial to all groups. Because the characteristics of the predominant group in the sample determine the results of this step (ordinarily called on item tryout), the test is usually sharply improved for that group (this is a desirable result) but relatively less improved for minority groups. The minority elements in the sample group do create noise in the data if they react to the materials in any way unlike the majority but this does not substantially affect the outcome. The characteristics of the majority group remain the determining factor in the process. The result is a better test for many children but a relatively more biased test for those minorities whose styles diverge from the majority of the tryout group. If the tryout group were predominantly black, blacks would be the majority group and the

would improve the tests more for them than others, i.e., it would tend to make the test biased against whites and other non-black groups.

WHAT TO DO ABOUT IT

A member of the testing industry, Green, D.R. (1972) suggests that racial, cultural, and ethnic bias in tests can be alleviated in tests by adding steps to the construction procedures used in building the norm-achievement tests and by using criterion-referred achievement tests for certain purposes.

Specifically, Green purports that "determination of bias must be an empirical procedure that includes direct examination of situation and data *after* materials have been prepared. The most promising solution to these dilemmas is to use item writers and editors that represent all major ethnic and cultural groups in the population, with each group producing a separate trial version of the test. The second step would be to try out *all* the materials on each subgroup *separately*. The third step would be to select items from all versions and edit them to best serve the interests of all groups.

Current achievement tests are focused on average performance, and their exercises are concentrated in a middle band of difficulty. According to Ralph W. Tyler (1970), "only about 5 percent of the exercise in commonly used tests are relevant to what is learned by students who are in the lowest third of their group. As a result we have not had means of evaluating the progress being made by the lowest third."

Tyler, as does Green, indicates that it is not impossible to construct valid exercises from both tails of the distribution, by obtaining the assistance of teachers and supervision and by careful revision in the light of tryout results.

A LOOK AT THREE SPECIFIC TESTS

As the raging debate continues over the "use" "abuse" "misuse" and "nonuse" of standardized tests it becomes increasingly necessary for secondary school English staff members to examine as carefully as possible the limitation of any standardized test that a district has chosen.[2]

A close look at the Iowa Tests of Basic Skills (Houghton Mifflin), Metropolitan Achievement Tests (Harcourt Brace Jovanovich) and the New York State Test in Reading-Beginning Grade 9 (New York State Education Department) indicates that the principal uses of these achievement tests are to (1) evaluate the status of a set of students in a class, in a school, or in a school system; (2) evaluate programs or projects, curricula, and instructional materials; (3) diagnose pupil, class, programs, or system problems; and (4) provide a basis for planning individual, class, or system programs.

Iowa Test of Basic Skills (ITBS)— Level 12, 13, 14, Secondary (1972 Edition)

This test is designed to test achievement in the language arts basic skills areas of vocabulary, reading comprehension and the language skills of spelling, capitalization, punctuation, and usage.

The item spread in the reading comprehension sub-tests allow for more reliability of scores than the other sub-tests. In each of the sub-skills tested under *Recognizing and Understanding Implied Facts and Details; Discovering the Purpose of Main Ideas of the Paragraph or Selection* there are 16, 13, and 13 items respectively.

The organization and evaluation item gives the user less reliability, for these are too few. The more "alike" the score for the subgroup in the specific sub-test, the more consistently or accurately the test measures; the more cases there are in a subgroup of individuals who make different scores, the more suspect we should be that there are "chance" factors. One way to avoid "chance" is to provide a large pool of items.

It is precisely for this reason that the English Staff at Isaac E. Young Junior High School has chosen to utilize, for grade diagnostic purposes, only the vocabulary, comprehension, and spelling sub-tests of the ITBS.

In the sub-tests of capitalization and punctuation all levels provide only one or two items per each discreet skill. Our staff feels that this simply does not yield enough information and that the chance factor is too high.

Houghton Mifflin, publisher of the ITBS, states that selections in the comprehension sub-test "were chosen in an attempt to represent as completely as possible all types of reading materials encountered by pupils in their everyday reading." In level 14 (grade 8 and 9) there are 10 items based on an ecology passage, 10 items based on an animal conservation passage, 8 items based on geography passage, 2 items based on a social studies passage (the selection deals with food process), 11 items based on a biology passage, 8 items based on a passage dealing with the evolution of the alphabet, and 6 items based on *one* poem.

Such examination of passages used to "test" skills in reading comprehension clearly points up the need to have school systems understand that comprehension is the province of *all* subject teachers, *not just* English teachers. Moffett (1973) points out that the skills "recalling," "relating implied and stated facts," "making inferences," "drawing conclusions," "interpreting and predicting outcomes," are not verbal or linguistic skills but *thinking operations*. He proposes further that "reading comprehension is merely comprehension."

Ronald Warahaugh makes a similar point in his book, *Reading: A Linguistic Perspective*... "When a person reads he is processing information".... It is the view of one of our reading specialists that visual, auditory, and linguistic fluency depend on the above skills. As the

"general" properties of thinking are sharpened, so is the potential for a "good" reader.

Other subject area teachers at the secondary level should be guided to develop subject-related concepts and skills simultaneously. Herbert Herber (1970) states, "There is a need for whole new strategies in teaching reading through content areas; a strategy that uses what we know about the direct teaching of reading, but adopts that knowledge to fit the structure and responsibilities for the total curriculum in each content area."

Our own item analysis on the ITBS indicates that students across percentiles did not do well on all questions that called for a complex set of skills in a given sub-test designed to test discreet skills. Kenneth Goodman (1973) calls this distortion of task. The following passage is an example from the ITBS level 13, grade 7. Here the student, in addition to providing answers for skills in the area of comprehension, must also compute.

> The blue whale, the largest living mammal, may soon be gone from the sea. These enormous creatures, that may grow to a length of 90 feet, are great prizes for whalers. Oil and other products from these and other species of whales are very important to a number of countries. Since World War II, the whalers have become so good at taking a catch that there may soon be no whales to catch. Experts say that there were 100,000 blue whales in Artic and Antarctic waters 30 years ago. Now they put the number at about 1000.
>
> An organization, the International Whaling Commission, is trying to save the blue whale from extinction. Its members, from fifteen nations, are trying to get their nations to limit the number of whales a whaling expedition may catch in any given year, particularly blue whales. However, this is a voluntary organization which has no authority to inspect the catch of a whaling expedition; so the slaughter goes on. A workable law is needed to get the whalers to limit their take. Members of the organization fear that such a law will not come in time and that the great blue whale is doomed.

Question 110. What is told about the whalers since World War II?

1). They have become very efficient at catching whales.
2). They do not obey the law in catching whales.
3). They do not catch any more blue whales.
4). They have agreed to limit the number of blue whales they catch.

Measures recognition and understanding of important facts and details.[3]

Question 111. A full-grown blue whale would be most like which of these in length?

1). A fishing rod
2). A large house
3). A rowboat
4). Three city blocks

Measures recognition and understanding of implied facts and relationships.

Questions 112. Which of these words is different in meaning from the word extinction as used in the article?

1). Extinct 3). Existing
2). **Destruction** 4). Extinguished

Measures deduction of meaning of work or phrases from context.

Question 113. The estimate of the number of blue whales now alive is about what part of the number 30 years ago?

1). ½ 3). 1/50
2). ¼ 4). 1/100

Measures recognition and understanding.

The 1972 edition of ITBS attempts to address itself to many of the design criticisms which were raised earlier in this article.

This edition does facilitate individualized testing if a district wants to glean certain barometers of information about a discreet population; it facilitates individual student item analysis, classroom item analysis, and building item analysis. Its norming design seems to be slightly more representative of city, rural, and suburban populations than the earlier editions.

However, it is my view that distortion of tasks, word isolation testing, and irrelevant selection of materials still exist in the test.

Metropolitan Achievement Test: Advanced

The Metropolitan is designed to test word knowledge, reading, spelling, study skills (includes source of information, dictionary, and library skill tasks), parts of speech, and language (punctuation, capitalization, and usage).

The sub-tests in reading are designed similarly to those of the ITBS Reading Comprehension, the New York State Reading, and the Stanford Achievement Paragraph Meaning Test (Advanced). Students read a paragraph, then answer questions about it.

There is a wider spread of items in each sub-test on the Metropolitan (Advanced) than in each of the other 3 tests.

The same source of confusion in vocabulary measurement exists in this test as in others. Words are presented in isolation and students are directed to select the "best" synonym from a number of alternatives. This method does not reflect reading vocabulary skills as an individual applies them in a practical reading situation. To determine a student's ability to cope with new vocabulary in context would seem to be the only reason for measuring. Of what use is simply taking a measure of the word bank of a secondary student? Goodman (1968) points out that "words in isolation," are particularly hard to *read* because there are no grammatical clues for the sentence structure or meaning clues from the context to help identify the word meaning; yet many sub-tests deal with isolated words.

There are 50 items on spelling in the Advanced Form. Students are asked to recognize incorrectly spelled words.

In examining more closely the spelling sub-test, one finds that the author, Bixler, selected the most popular speller for investigating representative spelling words. He assigned a grade placement in 1956. G.P. assigned for each word was the average year in school at which the words were first taught. His 1970 G.P. ratings for the words are based on a continuation of books examined.

Bixler's 1970 G.P. ratings for the words are based on a combination of his 1956 rating and the "Speller Median" results. Words retained the same G.P. as in 1956 unless the textbook survey seemed to indicate good reason to change in 1956 rating.

The test publishers assure users that in choosing the spelling of the incorrect items on the Intermediate and Advanced levels, much care was taken to ensure that the errors were those commonly made in the spelling of words.

According to the publishers of the Metropolitan Achievement Test their design for Standardization was as follows:

> It was decided that the major index variables to be used in selecting and describing the standardization group would be the following:
>
> (a) a socioeconomic index composed of median family income and median years of schooling for persons over 24 years of age
> (b) size of city
> (c) geographic region
> (d) public vs. non-public schools
> (d) mental ability test scores.
>
> The socioeconomic index was considered to be the most important non-test index variable since it has the highest relationship with achievement test scores of any non-test variables which have been studied. Median family income and median years of schooling were weighted so that they each contributed about equally to the total variance.
>
> Each pupil in the Metropolitan standardization, as well as in all supplementary research programs, took the appropriate level of Otis-Lennon Mental Ability Test, Form J. These data provide another source of information on the characteristics of the Metropolitan sample. Furthermore, since it was anticipated that many schools would use both Otis-Lennon and Metropolitan tests, this joint administration of the tests in the standardization would serve as an important base for comparing performance on the two tests. For procedures used in standardizing *Otis-Lennon Mental Ability Tests* see the *Technical Handbook* for that series.
>
> It was decided that Metropolitan Achievement Tests should be standardized at two times in the academic year. This decision was based on two facts, First, schools in the United States are split about equally in terms of preference for beginning of year or end of year testing. Second, interpolated norms obviously do not yield the accuracy given by empirically determined norms. If a test is standardized in the fall and has interpolated spring norms, schools using the test in spring cannot use the spring norms with full confidence. The same type of problem arises when a test is standardized in the spring and gives interpolated norms for the fall. For Metropolitan, it was decided that the standardization should be conducted so that schools could use the test's norms with equal confidence at the beginning and end of the year. Such a standardization procedure has the added benefit of allowing schools to compare academic growth of local pupils with a national norm within a single academic year.
>
> Metropolitan was standardized in October in the fall and in April in the spring since these months seemed to be the ones when the vast majority of testing is done.

Surveys of testing practices indicate that so little testing is done at mid-year that separate standardization in, say, December, seemed unnecessary. The only exception to this rule was the standardization of the Primer battery in the middle of Grade 1. The grade placements at which each battery was standardized are as follows:

Primer:	K.7, 1.5	Elementary:	3.7, 4.1, 4.7
Primary I:	1.7, 2.1	Intermediate:	5.1, 5.7, 6.1, 6.7
Primary II:	2.7, 3.1	Advanced:	7.1, 7.7, 8.1, 8.7, 9.1

The Item Analysis research took place in 1968. Sample pupils used for the item analysis had an average I.Q. of 100 on the Otis-Lennon Mental Ability Test.[4]

The percent of pupils in Item Analysis sample from various ethnic categories is provided by the publisher as follows:

Ethnic Group:	Native Born White	Foreign Born White	Negro	Puerto R. or Mexican	Oriental	Other
% of IA Group:	73.2	1.1	12.0	11.8	1.5	.3

All test publishers provide information as to the school system which participates in prepublication research programs (sampling, item tryouts, etc.)[5]

Although the names of systems are not the same, there seems to be significantly little differences in the "types" of communities used between the ITBS and the Metropolitan in the Northeast area. Both publishers, for example, used New York City sampling in their prepublication program; both used another major city's parochial school population—one Syracuse; the other, Rochester.

The readability of the Metropolitan Reading Sub-test passage can be viewed through the assignment of "noun grade level." The publisher used Elly (1969). The advanced form of the Metropolitan for all three batteries reports a noun average readability level of 3.36.

Validity in *any* achievement test must be defined primarily in terms of *content* validity. A test has content validity if the test item adequately covers the curriculum area that the test is *supposed to evaluate*. Each school system, and possibly each school building, has to evaluate the test in light of its own school population.

New York State Test in Reading, Beginning Grade 9 (1972)

The State Manual describes the purpose of this test as follows:

> The New York State Test in Reading, Beginning Grade 9, is a standardized achievement test designed to provide information which will be helpful to school administrators in evaluating the general effectiveness of their reading program in grades six, seven and eight. This test will be administered at the beginning of ninth grade as part of the New York State Pupil Evaluation Program.

Norms were not available for this test until the fall of 1973. At the time of this writing definitive information on the new norming sampling was unavailable.

The Test in Reading is based on the reading program recommended for New York State schools. It is designed to measure reading comprehension and provides only a total score. The test consists of 11 reading selections, each of which is followed by a number of questions. It gives a measure of the pupil's ability to read a selection and to understand it. The questions test the ability to recognize the central thought of a selection, to answer questions based on specific statements, to make inferences about the content of a selection, and to discover the meaning of a word from its context.

The test contains 50 multiple choice questions and requires 50 minutes of testing time.

The original Grade 9 test was set up to identify pupils needing special attention, to discriminate among pupils with marginal skill. The new form of the test is designed to correlate with the Grade 3 and Grade 6 tests.

The readability based on the Flesch ease formula and the SMOG formula determines that there is a sequential difficulty in the 11 passages ranging from approximately 5.0 to 10.0 in grade level.

Of the three tests observed thus far, the New York State Reading Test has less distortion of tasks and fewer word isolation problems. The reading passages, however, can be characterized as being similar to those of the ITBS and the Metropolitan, thereby, raising the same concerns for selection of passages.

CONCLUSION

English staffs must recognize the *severe* limitations of standardized achievement tests for the purpose of assessing growth of English and reading curriculums.

These tests do not measure the student's ability to use the tools of language; to invent, expand, substitute and transform sentences or ideas; to respond to literature; to act as a clarifier of fact and opinion; to be a creator with words; to behave as a user of images, an arranger of new groupings in the use of language; or to generate language and literary alternatives. These tests measure cognitive skills, basically at the knowledge level. They do not measure progress.

Though there are severe limitations, English staffs cannot ignore some usefulness of the information.

Standardized tests *may* serve as a useful barometer, as *one* component of an evaluation model for a school district. They have less validity at the building level and even less at the classroom level.

A final note about the purpose of standardized testing is stated precisely by Ronald J. Samuda (1973):

> In the final analysis we need to look at purposes of testing. If testing is to serve a selective and sorting function, and if, indeed, psychometric technology is intended to

preserve an elite, then it follows that traditional procedures for measuring intelligence and scholastic aptitude, tied to a set of middle-class ethnocentric norms, will serve that function very well. However, if it is our purpose to serve the mass of citizens, and if it is our goal to make measurement more facilitative for the education of the poor, of the minority student, and of the atypical individual, then we will need to expand our research endeavours so that psychometric technology becomes the handmaiden of educational innovation in optimizing the individual's competence. Through the qualitative analysis of achievement and weaknesses, we can point the way towards the modification of patterns of instruction which will match the individual needs of individual students. It is therefore the hope of such a philosophy of testing to contribute to the achievement of optimal developmental and educational opportunity for all.

Isaac E. Young Junior High School
New Rochelle

---

[1] English and English, A Comprehensive Dictionary of Psychological and Psychoanalytic Terms.

[2] All three test publishers cited in this article give clear advice as to the limitations.

[3] Publisher's directions

[4] Metropolitan Achievement Tests: Special Reports. Test Department, Harcourt Brace, Jovanovich, Inc.

[5] Ibid.

BIBLIOGRAPHY

Bloom, B.S., Hastings, I. L., Madau, G. J. *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill, 1971.

Eash, M.D., Grouping: "What Have We Learned?" Educational Leadership, 1961, 18, 429-434.

Elly, W.B. "The Assessment of Readability by Frequency Counts" *Reading Research Quarterly* 1969, pp. 411-427.

Goldberg, M.L., Passou, A.H., and Justman, D. *The Effects of Ability Grouping*. New York, Teachers College, Columbia University, 1966.

Goodman, K.S. *The Psycho Linquistic Nature of the Reading Process*. Detroit: Wayne State University Press, 1968. pp. 13-26.

Green, D.R. and Droper, I.F. "Exploratory Studies of Bias in Achievement Tests." Paper presented at the meeting of the American Psychological Association, November, 1972.

Heather, Glen. *Organizing Schools Through the Dual Progress Plan*. Danville, Illinois: The Interstate, 1967.

Herber, H.L. *Teaching Reading in Content Areas*. Englewood Cliffs: Prentice Hall, 1971.

Moffett, J. *A Student Centered Language Arts Curriculum Grades K-12: A handbook for Teachers*. Boston: Houghton Mifflin, 1973.

Schwebel, M. *Who Can Be Educated*. New York: Grove Press, 1968.

Sumada, R.J "Racial Discrimination Through Mental Testing: A Social Critics' Point of View." IRDC Bulletin, May - 1973, No. 42.

The Regents Comprehensive Examination in English: Some New Questions and Some Tentative Answers

by Lee Odell

Coming as a "final measure" of students' ability to think and compose, the Regents Comprehensive Examination in English ought to provide us with as much insight as possible into students' mastery of the processes of composition. Unfortunately, the exam does not do this. It cannot. As it is now constructed, the "composition" section of the exam forces us to ignore much of what practical experience and rhetorical theory tell us about the nature of the composing process. A close look at a Regents Exam essay, cited in "Suggestions on the Rating of Regents Examination Papers in English" as an example of superior work, should help explain the predicament.

> Gadgets! What a clever, yet all-inclusive word this one is. When we hear it we are reminded of devices which affect almost every phase of modern living and which range from the lowly jackknife to the highly sensitive electric eye. It is only after careful reflection that we realize how much actual influence these devices have on our daily lives.
>
> First we should appreciate the wide diversification of interests which gadgets serve. During the course of the day, we encounter gadgets at every turn. In the kitchen we find new-type can openers, apple corers, and cabbage choppers. The refrigerator has not escaped the inventive urge of our gadget scientists; lights to illuminate its interior, and plastic ice cube holders make the housewife's tasks easier. Our cars are distinguished by automatic direction indicators, window wipers, suction-type coat hangers. Truly we have come to regard gadgets almost as a part of up-to-date living.
>
> It is this "up-to-date-ness" that is another form of the "Keep Up With The Joneses" spirit of many Americans. The public's devotion to many fads and whims is now supplemented by the provocative appeal of gadgets of every description. It is sometimes their practical appeal which captivates the car buyer who admires the many devices which make driving easier; more often it is his desire to proudly display the latest in gadgets. Here is the complacency of a Babbitt at its worst.
>
> Is it not true too that gadgets result in our dependence upon them? We feel lost without them, and that something is missing from the completeness of our day. Who has not felt a sense of inadequacy when he finds that one of his favorite gadgets is not functioning properly?
>
> I realize my point of view is not a usual one, but let the reader ask himself as I have: Are gadgets really as insignificant as we think? I wonder.

The persona created in "Gadgets" is that of a polite, earnest but rather anonymous college-bound student writing for an audience of polite, college-educated adults who—at least for purposes of this exam—value writing that displays a particular set of stylistic features. There are no misspellings that would indicate he spells phonetically, lacks knowledge of spelling rules observed by many of his audience, or is letting his mind wander and failing to proofread his work. In addition, the speaker combines sentences in rather complex ways; he joins clauses with colons and semi-colons, and his sentences contain a fairly large number of embedded clauses. He chooses words that are relatively infrequent in teenagers' casual conversation; e.g., he refers to a "wide diversification of interests" rather than to "lots of different interests." In short, his writing exhibits many of the characteristics English teachers are likely to value. Moreover, the persona seems mildly provocative; he asserts a point of view that, he claims, not everyone accepts, but the questions he raises certainly do not challenge his audience's basic attitudes or values. He is argumentative without being abrasive.

Unquestionably, the writer has created a persona that is appropriate for his audience. Using thirty points as the highest possible score, judges indicated they would award the essay twenty-eight points if written by a junior, or twenty-seven points if written by a senior. Furthermore, the essay is clearly superior, in some respects, *to other* sample essays that accompanied it. Given the limitations imposed by the exam, the student has probably done as much as we have any right to expect. Yet the essay presents several serious problems, problems that have more to do with the nature of the "composition" section of the exam than with the quality of this specific essay.

Limitations of the Present Examination

For one thing, the exam precludes our determining whether stylistic features of the essay reflect a series of conscious choices or simply the rote application of stylistic formulae. I would like to believe the former is true. But my experience teaching composition to high school and university students does not provide much basis for such an assumption. Students can be incredibly inflexible about matters of style. Once they have learned that a given stylistic feature is effective in some contexts, they often assume that the feature is inherently good, appropriate in all contexts. For such students, writing an essay simply means following a limited set of clearly defined rules. On the basis of writing elicited by the present Regents Exam, we have no way to decide whether or not the writer of "Gadgets" is one of these students.

A second problem, one which perhaps causes the first, is that we have no way of knowing whether the writer of "Gadgets" can create other personae or appeal to other kinds of audiences. We do know that this writer can have a desired effect on an audience of English teachers. But we also know that once college students finish freshman composition, comparatively few of them (and fewer each year) spend much time writing essays to be read by English teachers. Rather, they begin preparing for careers that will require them to communicate with people whose backgrounds, values, and interests vary widely. Their practical experience in trying to communicate with these diverse audiences tells them what rhetorical theory and training should have told them all along: To compose effectively one must be able to create a variety of personae that will be appropriate for the various audiences one encounters. Sophisticated writers are those who can make difficult stylistic choices based not on a simple standard of "correct" and "incorrect," but on an understanding of the complex relationship between the persona they wish to create and the audience they wish to address. The student who wrote "Gadgets" may have achieved this sophistication. But the exam does not ask him to do the sort of work that will let us find out whether he has or not.

Another difficulty is that we have no idea of this student's ability to compose with non-verbal language. We can't tell whether he can choose the movement, posture gesture, and facial expression that will be consistent with the persona he's trying to project verbally. Nor can we determine whether he can use the body language that will help build rapport with, rather than alienate his audience.

Finally, "Gadgets" provides only a little information about the student's ability to engage in useful pre-writing processes. On the basis of the completed essay, we can make several inferences about the thought processes the writer used in examining his subject. He raised several questions, compared present day dependence on gadgets with that of a fictional character, and considered several causes of people's interest in gadgets. Since the time allotted for the exam had to be divided among examining his subject, writing a draft, and editing, we have no way of knowing whether the writer could have used these processes more thoroughly. (That is, could he have raised other interesting questions regarding the topic? Could he have thought of other significant causes for people's dependence on gadgets?) Nor do we know whether the writer had sufficient conscious mastery of these processes to use them in examining the complex, problematic materials he will have to deal with in college and in his career.

Toward a Revised Understanding of the Process of Composition

Until comparatively recently, it would have been difficult to argue that these problems do in fact exist. Most popular composition texts published over the last fifty years have led us to equate effective composition with adherence to what one such text calls "practical prescriptions for good writing." These texts, epitomized by Strunk and White's *The Elements of Style*[2] and Warriner's *Handbook*,[3] make no reference to non-verbal language and provide almost no insight into processes of pre-writing. At best, these works give useful instructions for creating one kind of persona. At worst, these texts offer only what I.A. Richards has called "the usual postcard's worth of crude common sense."[4] Fortunately, the extremely limited understanding of composition implied in these texts has been expanded during the past several years. A number of writers have enlarged our notion of the process of composition to include matters that we have conventionally ignored or relegated to other disciplines.

Speaker and audience: In *Points of View*[5] James Moffett and Kenneth R. McElheney describe different speaker-audience relationships that may exist in narrative prose. Selections in the text show how usage, diction, and syntax change as one moves from addressing an immediate, sympathetic audience to communicating with a remote audience whose feelings, ideas, and experiences may be quite different from one's own. In *Persona*[6] Walker Gibson examines the different voices that may exist in apparently "objective," formal, nonfiction prose. He points out, for example, that even such a "staid and responsible" publication as the New York *Times* may exhibit several different personae—ranging from the speaker in a news story, who cautiously reports only what "officials said" and tempers conclusions with "apparently" and "allegedly," to the authoritative editorial page speakers who pronounce value judgments and assert debatable conclusions, to the assured, informed, witty persona displayed in a feature or background story.

Neither Gibson nor Moffett and McElheney would argue that their ideas are without precedent. In fact, Gibson is especially careful to describe his indebtedness to Aristotle and classical rhetoric. Work done by Moffett and McElheney and by Gibson is important not simply because it may strike some people as completely new. Rather this work is important because it provides (in part, perhaps, reminds us of) rhetorical theory that allows us to go beyond the narrow view of composition suggested by Strunk and White and others. We can no longer equate "approximate perfection in technique of composition" with the ability to create a single kind of persona that is appropriate for only one kind of audience.

Our understanding of the process of composition has been further expanded by recent work with non-verbal language and by studies of pre-writing processes that help a writer examine new information, past experiences, personal values, etc., in an effort to formulate the ideas he will write about. We no longer need be limited by the assumption, implied in the present Regents Exam, that composing is chiefly a matter of drafting and editing, i.e., of writing out and revising one's ideas. As was the case in the argument based on Gibson's and Moffett's work, daily experience, as well as theory, argues for a change in our thinking about composition.

Non-verbal Language: We shouldn't need a social scientist to tell us that:

> Communication is a continuous interactive process made up of multi-leveled, overlapping, discontinuous segments of behavior. The interaction of communication does not cease when interactants lapse into silence, to begin again with the onset of phonation; other channels continue communication operations even when the auditory-aural channel is not in use. Humans move in relatively orderly fashion while

> they vocalize and when they are silent; they can perceive the regularity in the visible movement of others (or at least become aware when it is irregular) and proprioceptively in themselves. They can smell, taste, touch and otherwise register perception of themselves and their surroundings. When regularities appear, they are not simply mechanical, "automatic," or happenstantial. Research with visible body motion is convincing us that this behavior is as ordered and coded as is audible phonation. Like language, infra-communicational body motion is a structured system that varies from society to society and must be learned by the membership of a society if it is to interact successfully.[7]

Having to struggle with a social scientist's jargon is perhaps our punishment for having disregarded the substance of his message. The point is: effective, carefully considered use of non-verbal language is as important a part of communication as is careful choice of diction, syntax and usage.

A conventional response to this point is to acknowledge its validity but relegate non-verbal language to the realm of speaking rather than writing. Such a response implies a distinction between speech and writing, a distinction that does have some basis in fact. There are ways in which certain kinds of spoken language differ from certain kinds of written language. But an interest in *composition* (a term I think we should not equate with the more limited term *writing*) is not always well-served by this distinction. It obscures the fact that many kinds of communication which we often think of as "written" imply speaking and writing. Narration and dialogue imply people talking as well as writing—talking not only with their voice but with gesture, facial expression, physical appearance. Even analysis and exposition—in the form of lectures, reports, presentations—may exist in both a non-verbal and a verbal medium.

Most of us, I assume, have suffered through courses taught by professors who ignored this last point. These instructors' lectures—monotonously delivered by what seemed to be a disembodied voice that could cure the severest case of insomnia—should help support my argument by contrast. Certainly, the point is not lost on those whose livelihood depends upon their ability to persuade or influence others. Since polls indicate that politicians, lawyers, and advertisers—formerly good sources for illustrations of my point—now enjoy somewhat less public confidence than they once might have had, I'll give an example from a rather unlikely source. When the United Mine Workers recently sent men into the coal fields to encourage workers to accept a newly negotiated contract, they did not send bureaucrats, administrators, lawyers, or anybody who looked as though he earned his living by sitting around an office. Rather, TV newsfilm showed miners listening to rugged-looking types—solid, sturdy men whose gestures and mannerisms did not appear to be those of the carefully trained public speaker, men who wore open neck sport shirts, jackets (rather than sport coats) and slacks (not work pants, but not Brooks Brothers tailored slacks, either). Their performance accurately if not consciously, reflected

Aristotle's point: one important cause of persuasion is the "character" or "ethos" of the speaker.[8] In this day of visual media and reliance on face-to-face communication, part of that ethos is established by non-verbal language.

One may argue that students need to be able to compose a statement that is clear and effective even when not actually accompanied by non-verbal language. I agree. Students will frequently encounter situations in which this is important. But at least as frequently students will be expected to compose messages for a dramatic context in which non-verbal language will influence the meaning and the effectiveness of their message. In this latter situation it seems short sighted to argue that composing entails only selecting words and syntactical patterns. These verbal choices derive part of their significance from the non-verbal context in which they occur. If non-verbal language—e.g., that of the archetypal inept lecturer—subverts or obscures the message intended through word choice and syntax, it is hard to argue that the sender of the message has achieved real skill in composing. Since many modes of communication—narration, dialogue, exposition—may entail non-verbal as well as verbal language, and since the quality of the non-verbal language may affect the quality and effectiveness of the verbal message, I think we have a very compelling argument for *including non*-verbal language in our teaching and our evaluation of composition.

Pre-writing: We also have good reason to think that pre-writing is an important, measureable part of the composing process. We would not, however, discover this reason in many of the composition texts (by Strunk and White and others) that have long been used in many classrooms. These texts tend to assume as the writer of one such text puts it: "the stylistic side of writing is, in fact, the only side that can be analyzed and learned."[9] Consequently, they have emphasized matters of usage, sentence structure, diction, and organization, and they have paid virtually no attention to the problems of how we examine information, clarify our ideas, and discover what it is we wish to say.

The limitation of these texts is made apparent by ancient and contemporary theory. In his *Rhetoric*, Aristotle acknowledges that highly specialized types of reasoning must be treated in disciplines other than rhetoric. Yet his work clearly indicates that Aristotle did not concern himself only with arrangement and style (concerns that have, in this century, preoccupied authors of many composition texts). *Rhetoric* contains a great deal of information about how one explores a subject and formulates the ideas he wishes to communicate. Within the past decade, this interest in the processes of discovery has been revived by a number of writers. Some, like Edward P. J. Corbett, have been guided by Aristotle's theory. Others, such as William J. J. Gordon and Richard E. Young and Frank Koen have drawn upon recent work in creativity and language study. All of these writers suggest ways we can do more than simply exhort students to think carefully before they begin to write. Without minimizing the importance of non-conscious, intuitive processes, these writers have shown us how we can help students direct their thinking by creating "essential definitions,"[10] seeking analogies,[11] and making conscious use of certain intellectual processes.[12] Whether we are asking students to examine complex new information or re-examine familiar topics, we can help students master procedures that will let them do more than just sit and wait for inspiration.

Toward Revising the Exam

As we revise our understanding of the nature and processes of composition, we increase the number of interesting composing activities we can have our students do. But we also increase the number of problems we have to solve when we try to measure students' growth in composition. The present Regents Exam in Composition requires us to deal with one basic question: How do we measure students' ability to write out and edit a statement in which they create a persona that is appropriate for one kind of audience? Theory described thus far forces us to confront difficult new questions: How do we measure students' ability to make stylistic choices appropriate to a variety of speaker-audience relationships? How do we design an exam that will help us find out how well students can compose with non-verbal language? How do we get at intellectual activity that goes on during the pre-writing process?

In part, these questions are intimidating simply because they are comparatively new. Whatever its limitations, the present exam presents problems that are familiar; we at least have the satisfaction of knowing what sort of frustrations we will encounter when we think about the exam. New questions imply a new or substantially revised exam, and this, in turn, implies a new set of frustrations for us to deal with. This prospect is almost enough to make us content with the very limited notion of composition implied in present exams.

This contentment, however, comes at a very high price. It requires that we forfeit a reasonable understanding of students' ability to compose. Moreover, a desire for this contentment may lead us to ignore the fact that classroom practices, combined with the theory already described, begin to suggest some useful ways to answer these new questions. For example: It is fairly common to find English teachers assigning two different kinds of composition. In one, students are asked to analyze what someone else has composed. (E.g., Discuss the tone or point of view in a given short story. Examine the ways an author creates his characters. Assess the validity of the argument developed in an editorial, ad, speech.) In a second kind of composition, students are asked to show their own ability to create a persona, appeal to an audience, and examine some subject other than a literary work.

The distinction between these two kinds of assignments is not, of course, absolute. Nor is it adequately reflected in the customary distinction between creative and analytical writing. Any writing may be creative in that it reflects an attempt to come to a new understanding of one's self, one's audience, or one's subject. These reservations notwithstanding, the distinction seems useful. Both kinds of composing are important. And the ability to do one sort of composing does not necessarily imply the ability to do the other sort. Consequently, this distinction will help shape answers to the following questions.

How Can We Measure Students' Ability To Create Different Personae and Appeal To Different Audiences?

A partial answer to this question may be suggested in an assignment one teacher uses with eleventh and twelfth grade students. He asks students to choose from a list of traumatic experiences (e.g., you put a dent in your folks' car; you're accused of cheating) and do two pieces of writing about that experience. In the first, students are to describe the experience as they would in talking to a close friend who they know will sympathize with them. In the second piece of writing, students are to talk about the experience as they would to a more remote, less sympathetic, authoritarian figure. Students' interest in and success with this assignment suggests that the exercise might serve as a model for one sort of composition test item. In place of simply posing a topic such as amnesty for draft dodgers, we might ask students to take a position on amnesty and present that position to, say, the VFW and to a group of antimilitary college students. In such a case we might ask students to explain the persona they wish to create and to describe their assumptions about the values and attitudes they associate with each audience. We might evaluate the two pieces of writing by determining whether word choice, syntax, usage, and content are appropriate to the speaker/audience relationships students have described.

Another, more analytic format is suggested by work Stephen Dunning has done with ninth graders.[13] In an effort to get students to examine poetry more closely, Dunning has given students three versions of the same poem—the published version and two revisions in which he has altered diction, syntax, and metaphor. By comparing and contrasting specific features of the three poems, students are to determine which poem is the original. Using Dunning's basic idea, we might present students with two or more versions of the same prose passage. After describing the persona created in and the audience appealed to by the original passage, we could ask students to identify the original passage and cite specific features that helped them identify the original version.

To combine analytic and creative work, we might provide students with a description of the characteristics of a speaker and the audience to be addressed. We might then ask students to examine a piece of writing that does not reflect the persona we have described and is not appropriate for the intended audience. The students' job might be, first, to identify those features that are inappropriate for the specified speaker and audience, and, second, to revise the piece so that it creates the intended persona and is appropriate for the audience we have described.

How Do We Measure Students' Use of Non-verbal Language?

In answering this question, it seems useful to relate non-verbal language to considerations of speaker and audience. For an analytic test item, we might select photos (from advertisements or from a collection of photographs such as *The Family of Man*) in which a number of non-verbal clues suggest a person's state of mind, attitudes, personality. Students could be asked to list the non-verbal qualities that allow some understanding of the person depicted in the photo. To combine analytic and creative work, we could present students with a picture in which non-verbal language creates a persona that is appropriate for a given audience. We might ask students, first, to identify the non-verbal detail that helps make the persona effective for the intended audience and, second, to list non-verbal detail that would modify the persona so that it would be attractive to a different audience. In evaluating students' work on either of these test items, we might score students' answers on the basis of the specificity of their answers, their ability to cite more than one kind of detail, and the relevance of specific details to the intended speaker and audience.

How Can We Measure Pre-writing?

Different pre-writing procedures may require different ways of testing. Rather than trying to devise ways to measure specialized aspects of a single set of discovery procedures, it seems more useful to suggest ways to examine students' mastery of more readily accessible aspects of several pre-writing systems.

Creating "Essential Definitions": In his *Classical Rhetoric for the Modern Student,* Corbett explains Aristotle's procedure which "designates that which makes a thing what it is and distinguishes that thing from all other things." (p. 40) In dealing with a topic such as "gadgets," Corbett would tell a student to identify a specific gadget and 1) specify what this thing has in common with all other things that belong to the class "gadgets" and 2) specify the details that set this thing apart from other members of its class. An examination item based on this procedure might ask students to list the details that their subject has in common with other members of its class of things and then to list details that make it unique among members of its class.

Creating Analogies: Although Gordon did not specifically develop Synectics as a pre-writing procedure, one of Synectics' basic activities—looking for unusual analogies or metaphors—is obviously a useful means of clarifying one's feelings about and understanding of a subject. An exam item might ask students to begin examining a subject by creating as many metaphors as possible. A student dealing with "gadgets," for example, might identify individual gadgets and compare them with animate objects. Or that student might consider sub-topics such as people's need for gadgets, people's use/misuse of gadgets, the limitations of gadgets' usefulness. The student might then list all the ways he can think of that, for example, people's misuse of gadgets parallels their treatment of other objects or of people.

Posing Questions: Among other writers, Richard E. Young and Frank Koen argue that the investigation of a subject begins with one's awareness of a dissonance, a sense something is amiss or not quite right. Since this dissonance may be expressed as a question, Young and Koen argue that one's ability to explore a subject thoroughly may hinge on one's ability to pose and re-formulate questions. They argue that an essay consists of one's attempt to explore and answer the questions he has posed. Agreement with this premise might lead us to ask students to preface their writing by listing as many questions as possible concerning the topic they're examining. After a specified time, we might ask students to look over the list and eliminate or re-formulate questions that do not seem productive. They might subsequently write essays in which they attempt to answer the question(s) they have formulated. As an alternative to this procedure, we might ask students to examine a piece of prose in which a writer examines a topic rather superficially. We might then have students list significant questions which the writer appears not to have considered and write their own essay on the basis of questions not answered in the original piece.

Listing details or questions or analogies is, of course, only one part of the process we want to investigate. Thus it seems inevitable that a list of items would be followed by a short written passage in which students showed the results of their attempts to explore a subject. Given the limitations of the students' time and the careful focus of our interest at this point of the exam, it seems unfair to consider matters having to do with speaker and audience. We should probably be well advised to overlook features of diction, syntax, etc., unless those features obscured students' efforts to explain their ideas. In reading the essays, we should probably concern ourselves with one basic question: Have students been able to select from all the information turned up in the pre-writing process those items that will let them present a coherent, thoughtful statement? In judging the lists of items, we might consider two additional questions: 1) How many items does each student list? 2) Can the student list more than one kind of item? Quantity and diversity of items are not, of course, sufficient to insure a perceptive essay. They are, however, necessary and worth considering in our evaluation.

Learning from a "No" Answer

In one of his early books, John Holt tells of playing Twenty Questions with elementary school children. Eventually Holt's students learned to replace wild guessing with systematic questioning that sought to narrow the field of inquiry by eliminating certain possible answers. Confronted with the problem of trying to identify a specific number between one and 10,000, they learned that "Is it twenty-three?" is a less useful initial question than "Is it between one and 5,000?" Although pleased with this progress, Holt reports another problem at least as serious as complete reliance on guesswork. Holt says that if he responded to the latter question by saying "Yes, the number is between one and 5,000," students would cheer. But if he said "No," students would groan in dismay. Although each answer is equally informative, his students "still cling stubbornly to the idea that the only good answer is a *yes* answer . . . they have not learned how to learn from a mistake, or even that learning from a mistake is possible."[14]

Although the questions we are dealing with are infinitely more complex than those Holt posed to his elementary school students, Holt's story seems uncomfortably pertinent. If we try to pursue these questions, it seems only fair to assume that we are beginning a process which will, at least initially, be characterized by mistakes and unexpected, unsatisfactory answers. In her criticism of the Regents Exams, Patricia E. Speyser argues: "Anyone who criticizes an institution is likely to find himself criticized for failure to be constructive unless he also makes suggestions for its improvement, but I leave that work to those who have less confidence in the intelligence and competence of English teachers than I have."[15] In essence, I agree with Speyser's assumption. English teachers do have the intelligence and competence to make the needed improvements. But these changes will come slowly, perhaps painfully, and only with the combined effort of classroom teachers and researchers who realize they have learned something when they find out conclusively that the answer they are seeking is not between one and 5,000. It may be that answers suggested in this paper fall into this category. In any case, I look forward to the next effort that perhaps will tell us that the answer is not lower than 2,500 or higher than 7,500.

[1] Sheridan Baker, **The Practical Stylist**, 62nd edition (New York, 1959), p. 141.

[2] William Strunk, Jr., and E. B. White, **The Elements of Style** (New York, 1959).

[3] John E. Warriner and Frances Griffith, **English Grammar and Composition: A Complete Handbook** (New York, 1957).

[4] I. A. Richards, **The Philosophy of Rhetoric** (New York, 1965), p. 8.

[5] James Moffett and Kenneth R. McElheny, eds., **Points of View** (New York, 1966).

[6] Walker Gibson, **Persona: A Style Study for Readers and Writers** (New York, 1969).

[7] Ray Birdwhistell, "Kinesics," **International Encyclopedia of the Social Sciences**, Volume 8 (New York, 1968), p. 300.

[8] **The Rhetoric of Aristotle**, trans. Lane Cooper (New York, 1960), p. 8.

[9] Baker, p. 1.

[10] Edward P. J. Corbett, **Classical Rhetoric for the Modern Student** (New York, 1965).

[11] William J. J. Gordon, **Synectics: The Development of Creative Capacity** (New York, 1961).

[12] Richard E. Young and Frank Koen, **The Tagmemic Discovery Procedure: An Evaluation of its Uses in the Teaching of Rhetoric**, University of Michigan, Grant No. EO-5238-71-116, National Endowment for the Humanities (July 1, 1973).

[13] Stephen Dunning, **Teaching Literature to Adolescents: Poetry** (New York, 1966).

[14] John Holt, **How Children Fail** (New York, 1964) pp. 33-34.

[15] Patricia E. Speyser, "Springes to Catch Woodcocks: The Comprehensive Examinations in English," **English Record**, XX, no. 2 (December 1969), pp. 69-70.

SUNY, Buffalo

A Proposal for a Test in English for Students in Grades 11-12

Robert Carruthers

OVERVIEW

Most teachers of English in grades 11-12 are familiar with the "typical" classroom test for students in these grades. The test has a format and style which has become traditional and, to some teachers and students at least, unimaginative. (This does not mean that the test is not valid or reliable, however.) The same sections—vocabulary, library usage, spelling, grammar, etc.—march by in predictable fashion. More important, for the purposes of this report at least, there seems to be little if any relationship between the various sections of the test (with the possible exception of vocabulary questions based upon passages of reading comprehension).

The article below proposes a way of integrating in a test various English language arts and skills in a meaningful way, in contexts which are meaningful to teachers and students alike.

It is proposed that an examination be built which will consist of two sections, each worth 50 credits. The first section will be constructed around a *theme* of interest and relevance to students. The second section will involve a *process* requiring the students to use English skills in a life activity.

For each section, it is assumed that it is based upon the course of study.

I. *Theme* Section

Let us consider the *theme* section first. One has only to examine current anthologies, paperbacks, and other text and multi-media/multi-level materials to become aware of themes which engage high school students. Typical themes are "Adventure," "Alienation," "Coping," "Friendship," "Making Decisions," "Science in your Future," "Who Am I?" and "Youth Today." Often the teaching of these themes involves the integration of experiences in reading, writing, speaking, and listening. Why should not the test on this theme reflect this integration?

In planning this section of the test, the teacher may well do the following:

1. Select a theme which is of interest to students.

2. Select content from various media: books, periodicals, photographs, pictures, recordings, charts, etc.; use multi-level materials as well.
3. Include questions testing skills in reading, writing, speaking, and listening where possible.
4. Review 1-3 above to assure that each question/part of this section is logical for the context in which it appears.

The example which appears below is for the theme "Youth Today."

During the examination, the following would take place, question by question:

1. The teacher announces that the theme of this section (Remember, there are *two* sections to the examination) is "Youth Today." The teacher distributes answer sheets to the students.

2. The teacher reads twice two articles to the students. One article is about a teen-aged boy named Dennis, an above-average high school senior who goes on to college. The second article is about a teen-aged girl, Cathy, a brilliant senior who graduates from high school and then goes to work as a mailperson.

On his answer sheet, the student answers 10 multiple-choice questions testing skills in listening (for main idea, for supporting details) to the two articles. 10 credits

3. (a) The student reads a 150 word passage about a teen-ager. He or she then answers 10 multiple-choice questions which test comprehension, awareness of figurative language, etc. 10 credits

(b) The student then writes a 75-100 word paragraph in which he or she compares the person in (a) with a person in a current literary selection he or she has read, in one of the following respects: appearance, enthusiasm, sense of purpose, self-knowledge, or other respect. The student must cite evidence from the passage (in (a)) and from the literary selection as well. 25 credits

4. The student is asked to indicate *one* specific reference tool which he would consult for information about "Youth Today" and to indicate the kind of information he would find in that tool. 5 credits

50 credits

II. *Process* Section

The *second* section (50 credits) of the test would involve a *process* of significance to the students taking the test. The word "process" here refers to those life activities which most students will engage in at one time or another. One of these activities is getting a job, for example.

The teacher would need to prepare this section by completing the steps indicated below, and in this order:

1. Select the activity.
2. List all steps, in chronological order, involved in carrying out the activity.
3. For each step, identify specific English skills which it involves.
4. Decide which of these specific skills can be tested.
5. Define the objective of the skill to be tested.
6. Decide upon the type of question (short-answer, essay) or vehicle (tape, interview etc.) which will best test the specific objective.
7. Prepare, and edit carefully, the questions.

A brief outline of an illustrative activity and the skills involved in it appears below.

Getting a Job

1. Using available sources of information: a brochure on a specific job(s)	Reading comprehension and other basic reading skills	10 credits
2. Reading an advertisement	Reading; business awareness; understanding of abbreviations and commonly used terms	10 credits
3. Writing a letter of application (or resume)	Writing: expository, persuasive material. Observing form, content, and language conventions.	15 credits
4. Being interviewed	Listening: listening to a tape of an interview; criteria determined in advance. N.B. a speaking question, involving the interview situation, could be administered in advance of the test.	15 credits
		50 credits

III. *Advantages*

The two sections of the above test have several advantages.

For *teachers*, the test would reinforce the interrelationship and integration of the English language arts and skills. The test would probably have potentially greater validity and reliability than instruments presently used. Furthermore, the test might prove more quickly responsive to change than other instruments.

For *students*, the test would involve challenging experiences based upon the students' vicarious and real interests. The students would be encouraged by the attractiveness and relevance of the new test, and would see that it is related to their perception of the world. Finally, the test would give the students an additional reason for integrating their skills in English with content.

IV. *Cautions*

Care must be taken that the *theme* section be based upon a concept or idea which is broad enough in scope, which will yield many interrelationships, and which will appeal to both boys and girls. Also, each question should appear real and significant to the students.

For the *process* section, the activity selected by the teacher must be relevant to most if not all students. Each step must be defensible from the student's as well as the teacher's point of view. The artificial or contrived must be shunned.

For both sections together, a sense of balance must pertain. The generalized, abstract theme of the *theme* section can be balanced by a practical problem in the *process* section. Also, the test as a whole must present an adequate sample of course content.

The article above reports some of the experimentation presently being conducted by the Bureau of English Education of the State Education Department.

Please notify the author of your reactions to this article.—ED.

State Education Department
Albany

The Cooperative Development of Growth Measures for the Ability to Read Fiction

Thomas F. Callaghan

Introduction

In view of today's increasing emphasis on educational accountability, English teachers should re-examine some of their techniques for measuring growth in English skills. Although there are many areas of English instruction that are not amendable to multiple choice tests as growth measures, in some areas they may be quite useful. However, professionally prepared tests are generally not valid in measuring growth in certain specific skills in an individual English curriculum; and individual, teacher-made tests often lack the sophistication needed to make them highly reliable and valid. Yet, if teachers work together on the construction of multiple choice tests to measure growth based on very specific objectives, a reasonable amount of validity and reliability can be attained without an inordinate amount of individual effort. With certain specific course objectives as illustrations, I will suggest a method of cooperatively developing valid and reliable multiple-choice tests.

Test Design and Purposes

The tests employ a commonly used format and consist of a series of readings from lyric poems and narrative fiction with multiple choice questions after each sample. The basic purpose of these tests is to measure to what extent both the class and the individual student have progressed in specific skills during the course of a school year. Thus, content validity demands that the test items examine the specific skills mentioned in the course objectives for that year. Grades should not be given on the basis of these tests alone, for they obviously measure only a part of the entire course. Nor are these tests necessarily valid indicators of future performance. Their major purpose is to measure growth, although they may also be used to diagnose areas of weakness as well.

In a typical high school such tests could be given at the beginning and the end of each grade by the members of the English department. The pre-test and post-test for each grade could be cooperatively developed by the English department, refined through item analysis and used for years with high reliability and validity. An examination of class mean scores will be a very strong indication of the amount of growth in specific skills, but individual assessments of growth have to be made more cautiously and in connection with other growth measures.

The reading selections must not be familiar to the students nor should the student be required to have a knowledge of particular literary terms, texts, authors, or other critical information; for then the tests would be measuring something other than the skills specified. Teachers should not attempt to teach for these tests except to acquaint the students with the format. Indeed, teachers probably cannot teach to these tests. Moreover, if the English course is actually teaching what it claims it is teaching, the results will be apparent on the test results without any attempt to teach to the test.

Skills of Fictional Art

Characterization

The first set of example objectives deals with skills of fictional art. The first objective is that *the student will develop the ability to perceive and interpret characterization in the short story*. Each reading selection briefly describes a character. There are two questions after each selection, one to see if the student perceives the character as he is presented and the other to see if the student can correctly infer or interpret something about the character. (The ideas for the tests of this series of objectives are based on Charles R. Cooper's *Diagnostic Tests of Specific Skills* from the evaluation sequence of *Responding: Ginn Interrelated Sequences in Literature*, Ginn & Co.: Lexington, Mass., 1973.) The following is an example of such a reading selection and test questions:

> Arnold drew his overalls and raveling gray sweater over his naked body. In the other narrow bed his brother Eugene went on sleeping, undisturbed by the alarm clock's rusty ring. Arnold, watching his brother, felt a peculiar dismay; he was nine, six years younger than Eugene, and in their waking hours it was he who was subordinate...
>
> from "The Stone Boy"
> by Gina Berriault

1. Which of the following best describes Arnold?
    a) irritated,
    b) pessimistic,
    c) happy,
    • d) slightly jealous

2. Which one of the following would Arnold be most likely to do next?
    a) Go downstairs and eat dinner.
    b) Smother his brother with the pillow.
    c) Shoot his brother.
    • d) Ignore his brother completely.

Tone and Point of View

The second objective is that *the student will develop the ability to perceive and interpret tone and point of view in short stories and lyric poetry*. The first question asks the student to identify the tone or point of view, i.e., the emotional effect of the reading selection. The second question asks the student to explain his choice by choosing a statement that best supports the answer he selected in the first question. Such a test might look like this:

> Hand in clasped hand and side pressed close to side,
> Silently stand some children of the poor,
> And shyly, hungry eyes half-turned aside,
> Observe the eater through the open door.
>
> Anonymous
> Translated from the Sanskrit
> by John Brough

1. Which of the following best describes what the speaker is doing?

   a) Describing the nobility of man,
   b) criticizing people who do not provide for themselves,
   c) describing the beauty of childhood,
   * d) comparing one kind of man unfavorably with others.

2. Which one of the following statements best supports your answer to the first question?

   a) This is a common theme in Sanskrit literature.
   * b) The contrast between the two types of people speaks for itself.
   c) The poem has a lot of meaning in a few lines.
   d) The elaborate rhyme scheme conveys the tone.

Imagery

The third objective is that *the student will develop the ability to perceive and interpret imagery in the lyric poem.* Again a reading selection is followed by two questions, the first requiring the student to identify what is actually being compared and the second requiring the student to determine what the comparison implies or signifies:

> The moon was a ghostly galleon tossed upon cloudy seas...
>
> from "The Highwayman"
> by Alfred Noyes

1. In this poem the moon is compared to:

   a) a bottle,
   * b) a ship.
   c) a ghost of a man,
   d) debris.

2. What kind of night is it?

   a) snowy,
   * b) partly cloudy,

c) rainy,
d) moonless.

Comprehension and Interpretation

The second set of example objectives deals with close reading skills. Most of the same criteria apply to the tests of these skills as to the skills mentioned previously. *Each student will demonstrate growth in the ability to comprehend and to interpret.* The format of these tests differs from those above in that several questions dealing with these skills could follow each reading selection, which would tend to be more lengthy. (Alan C. Purves, *Literature Education in Ten Countries,* John Wiley: New York, 1973.) Comprehension involves the main thoughts of the selections, attention to detail, and an awareness of context. Interpretation involves drawing inferences from the subject matter and the language of the selection. There are, of course, other close reading objectives but these will serve as illustrations:

> A certain Mrs. Pickney, sixty-six, died in London Christmas Day
> It took the R.S.P.C.A.
> less than a week
> to carry out her final will. *Kill.*
> How speedy probate comes in England.
>
> Two donkeys dead, three cats destroyed.
> Mr. Peter Hall-Patch, who fired the shots,
> told the evening papers
> *"let's face it*
> *donkeys and horses*
> *don't worry about tomorrow*
> *just about their next meal."*
> Not a very romantic quote
> but poetry is lost on animals
> though for an instant every dying thing
> understands the language of his bullet.
> May Mr. Hall-Patch find a death humane
> under horses' hoofs
> unloved in the rain.
>
> "Joy to the World"
> by Rod McKuen

Many comprehension and interpretation questions could be developed using this poem. The first, second, and fifth test comprehension and the third, fourth, and sixth questions test interpretation. I am sure that you can come up with better questions than these:

1. Which of the following statements is not mentioned in the first six lines?

a) Mrs. Pickney died on Christmas.
b) Her will was quickly carried out by the R.S.P.C.A.
• c) Mrs. Pickney died suddenly.
d) Her will ordered that her animals be killed.

2. Who was Mr. Peter Hall-Patch?

• a) an employee of the R.S.P.C.A.
b) a murderer
c) a reporter
d) a poet

3. What attitude does the speaker express in the last three lines?

a) joy,
• b) anger,
c) contentment,
d) fear.

4. The title of the poem is used primarily because

a) the action takes place at Christmas
• b) The use of the word *joy* is ironic and means the opposite of what it says.
c) the use of the word *joy* understates the feeling in the poem.
d) the meter of the poem is like the song.

5. What did Mr. Peter Hall-Patch mean in lines 10-13?

a) He hates animals.
b) Animals eat a lot.
c) Animals sometimes die.
• d) Animals have no feelings.

6. The word *humane* in line 18 is used

a) as an exaggeration,
b) as an understatement,
c) as an accurate statement.
• d) with the opposite meaning.

Developing the Tests

What are the mechanics of cooperatively constructing and evaluating such tests? An English department of ten teachers could develop a pre-test and post-test for the first set of example objectives, if each teacher selected six readings and wrote two questions for each one. Each of the three sub-tests would then contain ten readings with twenty questions for a total of sixty readings and one hundred and twenty questions for both pre-test and post-test. You could easily develop tests for all four grades for all of the objectives mentioned in this paper and any other areas suited to these measures during a brief summer workshop. Obviously, a larger faculty would make the task even less time-consuming.

Item Analysis

Once you have collected and collated the readings and questions, you should analyze each test question for difficulty and discrimination to ensure test reliability. Often particular test items will be too easy, too difficult, or too ambiguous. Therefore, after you have given the test once, preferably to a sample group, analyze the test items to make sure that they are in fact fair to the students. (A. C. Crocker, *Statistics for the Teacher or How to Put Figures in Their Place,* Penguin Books Ltd.: Harmondsworth, 1969.) Item analysis is clerical work that should be done by computers, teacher aids, or students. The first few times you use the test you will have to change some items, but eventually the test will become a highly reliable and valid measure of some of your school's objectives for the English program.

Item Discrimination

Item discrimination analysis examines how well the test questions discriminate between the more able and less able students. It compares the top fourth or third of the papers with the bottom fourth or third for each question. If the question is doing its job, more of the top students will get the question correct than those at the bottom. The formula is: $\frac{H-L}{N}$ where H = the number of times the high group answered the question correctly, and L = the number of times the low group answered the question correctly, and N = the number of students in *one* of the groups (each subgroup should be the same size.) The results will fall somewhere between -1 through 0 to +1. A question which discriminates perfectly will get a score of +1; a 0 indicates no discrimination at all; and a -1 indicates that the question favors the low group. Items which fall below +0.3 should not be used again.

Item Difficulty

Although tests often begin with easy questions to give the student confidence, most tests later include a few questions which will challenge even the most able students. Your own experience as a teacher is very useful in deciding the difficulty level of the questions, but a calculation of item difficulty can confirm or contradict your choice of question. The formula is: $\frac{RIGHT}{N}$ x 100% where Right is the number in the entire class who got the question right and N is the number of students in the entire class. The majority of questions should have a difficulty index of 40% to 60%; the higher the percentage, the easier the question.

Conclusion

The procedures outlined above may seem very time-consuming, as indeed they are if attempted alone. Yet, when many are responsible for small parts of the test, a highly reliable and valid test can be constructed without an inordinate amount of effort. Moreover, the results more than justify any small effort, for the entire English department now can measure with greater accuracy growth in some very important skills in its English curriculum. You will be able to give an account of yourself with confidence.

East Aurora High School
East Aurora, New York

Are Six Weeks Gone Already? Or An Alternative To Measuring Growth In Writing Ability

Kenneth Watson

Introduction

Many teachers of English are faced every six or seven weeks with demands to evaluate and assign grades to students. The demands come not only from principals and guidance counselors who want to keep school records current, but from parents who are interested in their son's or daughter's progress in school. In fact, it was at the insistence of parents in Buffalo that in many subject areas the report card marking period was shortened from ten weeks to six or seven weeks to provide more frequent reports.

Certain aspects of the English curriculum are more readily adaptable to the short marking periods. Vocabulary study can be done on a weekly or bi-weekly basis, for example. A play can be read in two, three or four weeks. However, how does one determine growth in writing ability in such a short period of time as six or seven weeks? Does a grade of seventy-five on one marking period really indicate any measurable improvement over a grade of seventy on a previous marking period?

Furthermore, if an English teacher must include drama tests, book reports, and various quiz grades along with writing grades all on one composite grade for the marking period, how does the parent, the administrator, the guidance counselor or even the student know in what areas he has improved and in what areas he has done worse? Perhaps a better solution is needed. The solution should have two purposes: (1) separate the measurement of growth in the ability to write from measurement of growth in other areas and (2) space out the measurement over a longer period, allowing for ample, measurable growth to occur during the interval. Such a system would permit the student and parent to see the student's growth in writing ability alone. Secondly, it permits the administration to measure the growth in writing ability over a long period of time. The student's growth could be seen and measured during his entire four years of high school.

One such scheme that fulfills the above purposes was developed by Dr. Paul B. Diederich. It first appeared in *The English Journal* (vol. 55, April 1966) in an article entitled "How to Measure Growth in Writing Ability." In the following pages I will first describe Diederich's scheme and then follow that by a discussion indicating why I think Diederich's scheme fulfills the purposes I have set forth for measurement of growth in the ability to write.

Testing Conditions

Writing for a school test may be defined as a total process involving how a person feels, reacts, thinks and believes. To sample such a broad range of affective behaviors, a teacher needs a lengthy piece of writing. Therefore, it is important that the student be given an entire period (forty to forty-five minutes) in order to plan and write his essay. Topics for the essay test are chosen by all teachers in the English department prior to the day of the test. A final list of topics is made up by the department chairman. The final list of topics is not seen by either the students or the teachers until the day of the exam.

The test is given on the same day to all students during their English classes. Once students have come into the room and taken their seats, they are given two sheets of paper. One sheet (or booklet) will be used to write the response to the essay topic. The other sheet, which may be only a half or quarter sheet of paper, will be used to assure anonymity. Anonymity is important in the scoring of the papers. On this small sheet of paper the student writes the following information: his name, grade level, the teacher's name, the period the class meets and a six-digit number he selects. This six-digit number is one that he chooses without telling his teacher. It is written on both the small sheet as well as on the large sheet which contains his essay response. It is the only identification on the large sheet. After the test has been scored, the large and small sheets will be matched up by number so that the student's response may then be identified. An example of the small identification sheet follows:

STUDENT IDENTIFICATION SHEET

Six-Digit Number ....

Name .... Teacher's Name ....

Grade Level .... Class (Period) .... Date ....

The student identification sheets are collected and kept hidden until the tests are scored. After the test, the test papers are collected. All student papers are collected in one location, such as the English department office. Here all the papers are randomly sorted to assure that no teacher knows which papers he is receiving. The most thorough and most fair way to randomly sort is to sort the papers numerically by the six-digit number at the top of each. Thus a 000-031 chosen by a student in one class might go next to a 000-046 chosen by a student in some other class. This assures that no teacher knows which student papers he will grade. (At the same time, someone else would sort all the student identification slips by number so that they can be easily matched up with the papers later.)

Rating

There are two readings of each student paper. Each reader takes a portion of the papers (one-fifth for example, if there are five teachers) and sits down with a grading slip for each paper. The grading slip, which is the same for all graders, lists eight criteria. An example of the grading slip follows:

READER'S ANALYTIC SCALE

Topic Number .... Reader Number ....
Paper 6-digit No. ....

	Low		Middle			
Ideas	2	4	6	8	10	
Organization	2	4	6	8	10	
Wording	1	2	3	4	5	
Flavor	1	2	3	4	5	....
Usage	1	2	3	4	5	
Punctuation	1	2	3	4	5	
Spelling	1	2	3	4	5	
Handwriting	1	2	3	4	5	....
					Sum	....

There are several reasons why this particular scale is chosen over other types. First, it is an analytic scale. This divides the different aspects of the writing process into individual components. Thus, the reader is not forced to choose between a high grade and a low grade for a paper with good ideas and organization, but with sloppy handwriting or a plethora of spelling errors. Secondly, certain aspects of writing to which teachers may react strongly are singled out, so that a reader may assign a poor rating on those one or two aspects, and then assign a better rating on some other aspect. For example, if you have had an intensive spelling unit for the past few weeks in your classes and feel your students should know how to spell, then if you come across a student's paper that is filled with spelling errors you feel are careless and stupid, you may react unfavorably to the entire paper. Regardless of whether the paper belongs to one of your students or not, you have already judged that paper to be a failure. By separating spelling (as only one of eight aspects on the analytic scale), you are able to assign a poor rating on that aspect, but then you can reread the paper to assign fairer, and probably higher, ratings on the other aspects of the student's writing. Thus the analytic scale serves to alleviate somewhat the natural biases so that they do not result in an unfair rating of the entire paper, which is sometimes made up of a mixture of good and bad points.

The Reader's Analytic Scale will result in a rating of ten (all low's) to fifty (all high's). Since each paper is read by two readers, the total possible range will be twice that—from twenty (all low's) to one hundred (all high's). It is important, however, that the second reader does not see the results of the rating made by the first reader. When the first reader is done grading his set of papers, he should keep all of the rating slips in numerical order in a safe place, until the second rating is done. Also, no marks, corrections or comments should be placed on the paper, and, naturally, the two readers should not discuss the papers.

After the readers have exchanged papers, and after each reader has read and filled out grading slips on two sets of papers, then all the rating slips are gathered and the student identification slips brought out of hiding. If all of the slips have been kept in numerical order, it will be no problem to match them up by number with the essay paper. Each student essay response will have one student identification slip (identifying the author) and two rating slips (one each from the two readers). These three slips of paper should be stapled to the student's essay sheet. One problem might arise—namely, if two readers' scores on a student paper vary widely. If the scores differ by over ten points (eleven or more), then a third reader, such as the department chairman, should read the essay and either raise or lower one of the scores so that the resulting discrepancy is less than ten points.

It may be useful, beforehand, for teachers to obtain some practice on the analytic scale in order to improve the reliability of their scores. Using some papers not related to the test, teachers could meet as a group at a department meeting, individually rate copies of the same paper, and then discuss their ratings. This practice will enable teachers to learn how to rate, to conform their rating so that vast discrepancies do not occur, and enable readers to do the actual rating more quickly.

Using The Results

Now, what do you do with all of this? I would suggest an idea borrowed from the art department. As many of you know, colleges looking for prospective freshmen art majors ask high school seniors to submit a portfolio of their best art work. Why couldn't such a portfolio be kept for the student's writing that he has done throughout four years of high school? Isn't written expression as important as artistic expression of brush, crayon or stencil? If we, as teachers of English, really do believe that English is essential in modern society and deserves the four years of study required in most high schools, then why aren't we requiring (or even suggesting to our students) that their current work done in high school might be a valuable record that might be useful later on?

In some schools a student folder is kept in the English classroom. The folder is normally kept just for one year, however, and then the papers inside are turned over to the student in June and never seen or used again. It might be better for the student to see his papers from one year to the next. He can then see his own growth. The essays can form a basis for the student to measure his own growth, though the essays should not be the only papers included. Perhaps, student-written short stories, plays, research papers, newspaper writing, and book reports can go in. Also, at the end of each year, the student could write out his own evaluation (not a grade!) of work he has accomplished during the year, including literature read in class, books read on his own outside of class, compositions written during the year, and other material the teacher or student wishes to include.

In the preceding paragraph I have tried to stress the importance of student writing, including the essays, because I feel as a teacher of English that it is important to have a complete, continuing record of student growth. This is the only way that real measurement of growth is possible.

To come back to the rating scheme: the individual student essay then is placed in a folder. To this is added three other essays written later in the year, so that the student has four essays, written approximately every two months (November, January, March and May). The four papers, each having been read by two people, form a composite of eight different scores on the student's writing. Since each score varies from ten to fifty points, the total for the year will vary from eighty to four hundred points. Research in written composition demonstrates that it takes at least four papers and eight readers to satisfy reliability criteria for the type of comparison which I am going to suggest next.

Student writing placed in a folder is very necessary for the student to see his own growth, but parents, guidance counselors and administrators are also interested in seeing how the student's performance compares with the performance of other students in his grade and in the school at large. For example, as a basis for admission to college many colleges now request information on a student's rank in class. Diederich's plan yields this type of information as a measure of growth in writing.

Since all the scores will fall between eighty (minimum) and four hundred points (maximum), a teacher can arrange all of the grades numerically from highest to lowest. Perhaps this can best be done by initially dividing the papers into several smaller, more manageable groups, perhaps with divisions of 150 and below, 151-200, 201-250, 251-300, 301-350, and 351-400. Then the papers can be further subdivided, until they are all ranked from the lowest (80) to the highest (400). Since this distribution will include all the scores by all the English students in the entire school, a student knows his standing in relation to the whole school. Presumably, many freshmen will be found closer to the bottom, and more seniors will be found nearer the top. This is as it should be. If it were otherwise, then English teachers should look for different jobs. If one were interested in finding a student's rank in class, he has only to count how many papers of students of his own grade level are above him. If the papers (now completely scored) are coded some way to indicate grade level this could be relatively easy. Perhaps a color for each grade level (blue=freshman, green=sophomore, red=junior, black=senior) marked in the upper right hand corner would aid identification. This would make it easy to rank a student in his class (for example, as the fifteenth highest freshman) as well as in the school at large (for example, as the one hundred twentieth). Thus the freshman student knows he did extremely well in comparison with other freshmen, but not as well as say the fifteenth ranked senior who perhaps scored thirty-seventh in relation to the whole student body. The student score and rank could be listed as follows:

Name	Rank in class	# students in class	Rank in school	# students in school
Freshman	15	(300)	120	(1200)
Sophomore				
Junior				
Senior				

This slip of paper should be kept in the student's permanent English folder, and each year his new rank in class and rank in school should be added. Thus a normative (or comparative) measure of growth in writing ability can be readily seen.

From this permanent record a teacher of English can see that a student's growth in writing ability can show progress in two ways—both in rank in class and also in rank in school. Of these two measures, perhaps the most important is growth in rank in school. Any upward movement in rank in school indicates real growth. In comparison, growth in rank in class means that a student is learning faster than his classmates, but growth in rank in school indicates that the student is assimilating skill during his four years in high school. A student may not change rank in his own class (grade level) very much at all during his four years. He may have ranks like the following: freshman (96), sophomore (102), junior (98), and senior (100). Does this mean the student has not grown during his four years? No. That same student may have had the following ranks in school: freshman (1050), sophomore (800), junior (525), and senior (100). The increase from rank 1050 (near the bottom of the school) to 200 (close to the top) indicates that the student has grown in his ability to write during his four years, a fact not evidenced by his rank in class alone.

Responses to Criticism

Certain variables on individual pieces of writing still present their problems in this type of test situation as in the normal classroom tests. For example, a student may feel upset, fatigued or for some other physiological or psychological reason may not be able to do his best on the test. Also, even with two readers for each test, both readers may grade their sets of papers harshly, resulting in significantly lower grades for the sets of papers. (A movement to the left of one column on the Reader's Analytical Scale sheet results in ten fewer points out of fifty possible on the score.) Thus instead of receiving 8-8-4-4-4-4-4-4- (forty total), a student might receive 6-6-3-3-3-3-3-3 (thirty total). Even teachers have their bad days. However, in light of these variables affecting students and teachers, isn't it better to have four papers from each student per year, and eight graders, rather than one final exam grade graded by one teacher? Even if exams are retained, the essays, along with the ranks in class and ranks in school form an excellent second look at the student's growth in writing ability.

Perhaps the principal criticism against undertaking this project is the amount of time involved. How long does it really take? Bear in mind that you make no red pencil marks on the essay. Furthermore, first impressions are valued. In keeping with this, the reader should spend no more than two minutes per paper—time enough to read about three hundred words. In scoring two papers, he will have spent a total of only four minutes rating both. On the average a teacher may spend from five to eight minutes per paper, marking mistakes in spelling and making comments on ideas or organization. So this rating of two sets of papers really takes less time than the normal time needed to grade a set of student compositions.

Furthermore, what do you have when you finish? Normally, after making many red-penciled marks you have only one more set of compositions to return to the students. You know how well Johnny did in comparison with other students in your room, but you can only guess how well Johnny did in relation to other students in his grade. Is a seventy-five in your class the same as a seventy-five from another teacher? Diederich's scheme yields each student a rank in his whole grade level and in his whole school, not merely in one teacher's classes.

What I am trying to show is that you as a teacher of English have something permanent and verifiable. You have something that has been evaluated and scored by other faculty members of your department. No parent, administrator or student can claim any unfair bias on your part in the scoring of the essays. Furthermore, you have verifiable proof of a student's growth in writing ability throughout four years of high school. At the end of four years a student will have written sixteen of these essays (four each year) and have been assigned four ranks in class and four ranks in school at large. Upward growth in the latter rank is proof that the student has learned to write better.

I feel that the above reasons strongly urge us as teachers of English to use Diederich's scheme to measure growth in writing ability. Only in this way can we prove to our critics that we do have real, measurable growth in writing ability occurring in our schools.

Riverside High School
Buffalo

Grades and Growth in the Writing Program

Charles R. Cooper

Introduction

After a student has produced a piece of writing, how do we help him answer these questions: What have I done? How good is it? What should I do next time? In this article I want to propose partial answers to those questions. The answers will involve us in what is increasingly referred to these days as *formative evaluation*, a process of response and feedback that informs the student of how well he is doing and informs the teacher about the effectiveness of instruction or student practice. Since this new term and others related to it are often used in connection with school district measurement and evaluation schemes and with teacher accountability, I want to spend some time discussing these terms in this introduction to the main part of the article. In particular, I want to show how some of the new measurement and evaluation concepts have limited use for English teachers.

The term *formative evaluation* is most often used in the context of *mastery learning*, an instructional scheme of some use in biology but probably of no use in writing. Conventionally used, *formative evaluation* implies (or requires) short, sequenced units of instruction which can be "mastered," the mastery being demonstrated by unit tests, usually criterion-referenced, that is, specifically written for the stated objectives of the unit. For example, the student studies the concept of *photosynthesis* for two weeks and then takes a mastery test; if he scores above 80%, he goes on to the next unit of study. Obviously, such an instructional-measurement approach is not useful to writing teachers. There are two main problems with it: (1) it assumes that points or plateaus of mastery can be identified and (2) it requires that the subject or skill be broken down into small, sequential, testable units.

Can a high school junior "master" the essay of argument or the autobiographical incident, or even the descriptive paragraph or the business letter? Can a college junior? Isn't writing a skill that one can only get better at but never master, except in superficial aspects of scribal accuracy (spelling) and format. I long ago mastered the *format* of the business letter, but I never think of myself as a masterful writer of business letters, of having reached some plateau of efficiency and notable high quality. Each new letter is a unique problem, requiring important judgments of tone, organization, diction, and syntax. Often my business letters literally fail: they irritate when I meant to soothe, or they alarm when I merely meant to inform. Therefore, from our own experience as writers we can conclude that the conventional concept of mastery learning is not useful—is probably even quite misleading—to us as writing teachers.

The other major problem with mastery learning as a way of thinking about teaching writing and evaluating growth in young writers is that it leads to fragmenting and false categories. What are the units of the activity of becoming a better writer or of the activity of composing a single piece of writing? Pieces of writing are composed of words, phrases, sentences, and sometimes of paragraphs; but those units do not provide a way of planning and sequencing writing instruction. Composing is a holistic activity. Once the writer has produced a piece, then there is much we can say to him about its language, syntax, rhetoric, style, and effect; but we cannot anticipate what a particular writer needs to know or practice until we have seen his writing. There is nothing about the writing process or about written pieces that will permit us to plan instruction in short units for all students to master, even at their own speed in a programmed format, with mastery demonstrated by tests. Writing has no content, in the way biology has content. Writing is a skill to be practiced. There are no incremental, sequential behavioral or performance specifications for such a skill. The writer writes and then sees what he has. He gets response from others. Then he revises or writes another piece. Writing instruction must enhance that process, not subvert it with trivial and indefensible mastery units.

There are certain syntactic patterns and word usage features students can usefully practice independent of the act of writing or revising. We know from research that a special kind of sentence-combining practice enhances written syntactic fluency and that oral-aural drill on particular usage problems can be quite effective, but this arhetorical practice cannot be viewed as preparation for writing—it should not fill up the first few weeks of class time before the student writes his first piece—and it cannot be tightly sequenced in small, testable units on any grounds I know of. Equally important, it cannot be tested in conventional mastery unit fashion. A valid "test" of syntactic maturity and control of the forms of standard usage can only come from examining the student's writing: at the end of the year is he making fewer usage deviances per 100 words and are his independent clauses longer on average, more deeply embedded, more varied in structure?

In some schools the notion of mastery learning and the idea of individualized instruction have conjoined to create a monster, sometimes named LAP (Learning Activity Packet) or PIK (Personalized Instruction Kit). Nearly all the writing LAP's I've seen are horrors, directing the student to work in isolation through some pages in a workbook or programmed textbook and then take a mastery test. Individualizing instruction is a fashionable (but important) development which we don't need to be anxious about. Good writing instruction is always individualized, even when the teacher gives the writing assignment. The student writes a piece which gets quite personalized attention from the teacher and from other students. And then perhaps the student revises; and that piece gets another personal response, not the predictable programmed instruction cajolery of the cathode ray tube or the workbook frame or the empty satisfaction of scoring 86% on an unhelpful LAP mastery test on naming sentence parts, but a sensitive human response which can take any of the forms in my formative evaluation scheme outlined below.

Let us admit, then, that writing is not a skill that high school and college students "master" and that it is not a skill or activity that can be fragmented into units for sequential instruction. Let us use the notion of *formative evaluation* to mean only response and feedback to the writer's efforts, feedback that describes for him what he has done, conveys unmistakably how other readers (the teacher and other students) understand what he has said, and indicates unambiguously what he must do differently when he revises or writes the next piece. This feedback would be concerned mainly with rhetorical and stylistic matters, and only secondarily with matters of correctness, with spelling, usage, and punctuation.

This article proposes ways to provide that feedback after a writer has produced a piece. In a sense it is an article about writing instruction, and yet I want to try to keep the focus more narrowly on formative feedback, omitting such important matters as guided prewriting activities and the teaching of stylistic refinements, intellectual strategies, and syntactic fluency, not to mention planning the writing program and sequencing writing tasks. I will also not deal directly with *summative evaluation*, the measurement of growth in writing ability over a period of time. For information about growth measurement see Kenneth Watson's article in this issue and my article in the March 1975 *English Journal* (Cooper, 1975).

Grading

I am also not concerned here with the question of assigning grades to pieces of student writing. Grading individual pieces subverts the formative evaluation process I want to describe below. *We have to give up putting grades on student writing*, and instead concentrate on providing in the classroom the response a writer must have to what he is producing. Grades do not even contribute to summative evaluation. A row of grades in the gradebook is not a measure of anything. *The writing teacher's task is to foster growth and to describe what growth is occurring.* Once we have described a writer's growth over a period of time (several weeks or months), then we can assign a public grade to it if we have to—and we still have to in most schools and colleges.

The last argument of the defender of grading every piece is that captive adolescents will not continue writing unless each piece is graded. The experience of many teachers now easily refutes that argument. Students will continue to produce large amounts of writing if the writing tasks are not uniformly dull and pedestrian and unrelated to their experience, if they can see some purpose for writing beyond pleasing the teacher, and if they can write for many audiences, not just for the teacher. Students *do* need response to each piece, and that is what the procedures I am going to outline below are designed to do.

This is an article, then, on what to do besides grade writing and give tests on information about writing and language. It is an article about formative evaluation English style.

Formative Evaluation for Guiding Growth in Writing

Formative evaluation—to summarize my use of the concept here—is helpful response to the writer's efforts. It tells him what he has done, how effective he has been in communicating with his intended audience, and what he might do differently next time he writes in that mode for that audience. It goes on continuously with each piece the writer produces, shaping his scribal activity, moving him slowly in the direction of what mature writers do, but not measuring him against an adult criterion, not expecting "mastery," but only some demonstration of growth from the point where he started in the course. It should point out mistakes, confusions, ambiguities, and failures; but it should be mainly on the positive side in order to encourage the writer to try again and again. It should foster confidence and reward freshness, vigor, and honesty. Most important, it should help the young writer find his own voice.

The teacher guides this evaluation process, participating in it himself, but also arranging for students to participate, training them to do so effectively. What exactly do the teacher and the students do? I want to propose four general activities.

1. *The teacher talks to the student or writes comments on his paper.* Making comments or corrections on the student's paper without the student actually present should be done only as a last resort or as an expedient. A conference with the student about his paper is always preferable. Even with classes as large as they are, we must still do as much conferencing as we can possibly arrange, even if it means devoting several days a month to it, with activities for the rest of the class members carefully planned so the time is not lost for them. We must also patiently but persistently continue making the case for smaller class sizes, even in these hard times. The new concern with "basic skills" may aid us here. Writing is a most crucial basic skill, and good basic skill instruction cannot be had on the cheap.

Whether we talk or write comments, we should initially withhold judgment about the worth of the paper as well as overlook usage, spelling, and transcription problems while we concentrate on (1) making a personal response to what the student is trying to say, to the composition behind the transcription and (2) describing to the student what he has done. Some papers, written for the teacher as trusted adult, will require only the personal response. Description, detailed comment, and editing are best used selectively on papers that have real promise, papers that may be published in some form for the class—read aloud by the student or the teacher, put up on the bulletin board, pasted into the class writing scrapbook, published in the class magazine, or sent off elsewhere to be published, perhaps in the school newspaper or literary magazine.

It is very helpful if the student keeps all of his writing in a folder or if he accumulates it along with revisions in a spiral notebook. At each conference he brings *all* of his work with him, not just the last piece. If students are keeping a journal along with the regular writing we are assigning, we can alternate conferences on the journal and on assigned writing.

Good conferences are intense, focused and fast-paced, the latter being important just to get around to everyone. The conference should always begin with the personal response of a sympathetic reader, the teacher, to a young writer. It can then move on to a description—not an assessment or a judgment—of what the writer has done. We can deal with matters of organization, syntax, and diction, all within the framework of the writer's purpose, subject, and audience for that piece. We probably should be selective, not trying to cure all ills at once. If the paper warrants it, we can make suggestions for revision; or we can just ask the student to try another piece in that mode. Gradually, the student can do more and more of the talking in the conferences. We can ask him what is the best thing in his paper, the weakest thing, what he would do with the paper. We can encourage him to explain and defend, to "say what he meant," to begin to take responsibility for his writing.

The best articles I have seen recently on responding to student writing are by James Klein, Lois McCallister, Donald Murray, and Stephen Judy. They are all listed in the bibliography to this article and I recommend them enthusiastically. Murray, a Pulitzer Prize winner and a writing teacher, has this to say about the teacher as reader of student writing: "He doesn't look for what he hopes will be there or ought to be there, instead he listens for what is there. He is ready for the unexpected, alert to a hint, a word, a tone, a fact, a shape, a subject which will make him respond as a reader, a human being, not an English teacher." Judy's extremely helpful and detailed article outlines seven steps in the process of responding: listening for the student's voice and deciding whether the piece is real or fake, responding in a personal way in a conference or in written comments, deciding whether the piece should be kept private or made public to the class, determining a form of publication or distribution in the class, editing, copy reading, and "publishing." Only a few pieces would make it all the way through this process.

A very promising possibility for response to student writing is outlined in an important recent article by my colleague Lee Odell (Odell, 1973), who is studying the intellectual strategies (for example: focus, contrast, change, reference to sequence, reference to physical context, classification) apparent in students' writing. Odell argues that "if we are to make genuinely useful responses to students' writing, we need to devote a good deal of our effort to identifying mental processes implicit in their language and helping students add to and refine strategies they already possess. . . . we can help them master some of the processes that are prerequisites to effective writing. . . . When we respond to students' writing by helping them enlarge and improve upon their repertoire of intellectual strategies, the usefulness of our response extends well beyond a specific assignment for a single course." There is some evidence students can be taught to increase the use of appropriate intellectual strategies in their writing (Odell, 1974).

2. *The teacher leads whole-class discussion of representative pieces of student writing.* Since we could argue that there is no such thing as a "representative" piece of student writing or that writers are really interested only by talk about their own pieces and not those of someone else, we should use this activity cautiously and sparingly. It should never become a substitute for the teacher-student conference. Carefully planned, however, it can be quite effective. We ditto a piece of student writing or we make an overhead projector transparency of a piece. Selecting a few things of interest in the piece, we lead a class discussion about it. Eventually, the students can do most of the talking. We have all done this sort of thing, but let me illustrate how it might look in the classroom. Here is a piece of writing by a twelfth grader:

My Best Class

[1]The best class I was in was my eleventh grade history class. [2]I found this class very interesting because the teacher looked upon each student as a young lady or young gentlemen and not as little children. [3]We studied out of a textbook, but the class was still interesting because she would relate whatever topic we were on to the current news. [4]She was always pleasant and willing to share some news article with the class. [5]She would choose a topic which she knew would interest the class. [6]The class was always in order. [7]As far as I can remember there was never a disturbance in the class. [8]She would encourage you to do your homework, but if you didn't do it she would tell you, "You are only hurting yourself and no one else." [9]She would never force you to do anything you didn't want to do. [10]She was a very good teacher.

Now let's assume we have chosen this "representative" piece from the class in order to point out some possibilities for increasing syntactic fluency and variety, being careful to emphasize the point that syntactic choices have to be made within the context of the whole rhetorical situation—someone saying something for a reason to a particular audience.

Although the mean length of the independent clauses is only 12.6 words for the thirteen independent clauses in the ten sentences in the piece (about that of the average eighth grader) the piece has a reasonable degree of syntactic fluency. Sentences 2, 3, 7, and 8 have adverb clauses. Sentences 5 and 9 have adjective clauses. Sentence 3 has a noun clause. And yet we notice right away the sentences are not very complex structurally, not very deeply embedded (except for Sentence 3 with the "whatever" noun clause inside the "because" adverb clause), and lacking absolute phrases, strong medial modification, or noun substitutes in the subject position.

What could we say to the class about this piece? We might suggest the following possibilities:

1. Begin Sentence 2 with the "because" clause.
2. Begin Sentence 3 with "Even though."
3. Embed the first part of Sentence 4, "She was always pleasant," in Sentence 2 as follows: ". . . the teacher was always pleasant and always looked upon each student . . ."
4. Then begin Sentence 4 with "Always willing" and attach it to Sentence 5 following.
5. Attach Sentence 7 to Sentence 6 by creating an absolute beginning "there never being." (Here the use of the term "absolute" would not be important. Students will see the usefulness of the new structure without ever having heard about "absolute phrases.")
6. Attach Sentence 9 to the first clause in Sentence 8 by creating another absolute beginning "never forcing."
7. Then make a new sentence of the second clause in Sentence 8 by beginning "If you didn't want to do your homework . . ."

Now with this short piece I don't want to make any grand claims for the overall effects of these changes. But I have demonstrated how elements can be moved around in a piece of writing to achieve greater structural variety. I did create a participle phrase sentence opener ("Always willing . . .") and two absolute phrases, and I increased the mean length of the independent clauses to 20.5 words, well beyond that of the average twelfth grader.

These teacher-led discussions can focus anywhere in the piece. They can either be planned to show the students something in particular or to draw them out about the piece, getting them to suggest changes. A great advantage of these discussions is that they are a model for the students of what can be said about a piece of writing, an insight they need to develop to work helpfully in the two remaining activities I want to discuss.

3. *In pairs students give each other guided feedback about their writing.* The activities I want to recommend could also be used in small groups of five or six, but I have found they work best with pairs. If one student writes comments and makes marks on a paper, the second student to look at the same paper is going to be unavoidably influenced by those marks and comments. Furthermore, one set of marks and comments may be enough for one writer to heed, particularly if the teacher gives feedback on the piece as well. Having several students read and respond to a piece is also very time-consuming. When two students exchange papers, they can get fairly far into comments and discussion on both papers within fifteen or twenty minutes.

I have found that at first these student pairs, even college students and graduates, need specific directions. They need to be told what to look for and what to do. They will need considerable help in order to get beyond bland judgmental reactions. In fact, the most useful activities I have seen *preclude* judgments. Eventually, of course, we want students to relish these read and respond sessions, to be confident in their responses, and to know how to give helpful, focused feedback.

The best activity I have seen—students from Grade 7 to graduate school like it and can do it easily—comes from Peter Elbow's book *Writing Without Teachers* (Elbow, 1973). It has three stages: *pointing, summarizing,* and *showing.* I will explain the first two here and encourage you to read Elbow for the third.

Pointing involves (1) putting a straight line under words or phrases that you like and (2) putting a wavy line under those you don't like. Elbow describes the first in this way: "the words and phrases which most successfully penetrated your skull: perhaps they seemed loud or full of voice; or they seemed to have a lot of energy; or they somehow rang true; or they carried special conviction." And the second this way: "words and phrases that strike you as particularly weak or empty. Somehow they ring false, hollow, plastic. They bounced ineffectually off your skull." The student *points* on his *second* reading of the piece, and the decisions are entirely his own. In discussion immediately following, each student reader can explain as well as he can to each student writer why he *pointed* as he did.

Summarizing involves these four steps (quoted directly from Elbow):

1. First tell very quickly what you found to be the main points, main feelings, or centers of gravity. Just sort of say what comes to mind for fifteen seconds, for example, "Let's see, very sad; the death seemed to be the main event; um. . . but the joke she told was very prominent; lots of clothes."
2. Then summarize it into a single sentence.
3. Then choose one word from the writing which best summarizes it.
4. Then choose a word that isn't in the writing to summarize it.

Make copies of these steps for the students, and encourage them to follow the steps exactly at first. *Summarizing* is to be done informally without too much planning or thinking. It shows the writer how his piece struck one other reader. As with *pointing,* discussion can follow immediately between the two students.

Another source of ideas for directed feedback is Ken Macrorie's *Telling Writing* (Macrorie, 1970). Scattered throughout the book are ideas like this one: to identify repetition draw a circle around all the repeated *words* (besides *a, and, the, of,* etc.) so the writer can consider whether he needs to keep all of them, and draw a circle around words and phrases you think repeat *meanings* unnecessarily. I highly recommend Macrorie's Chapter 9, Criticizing, for a general discussion of what it means to give and receive criticism of writing.

Pairs can help each other with any aspect of the writing they are doing. Sometimes students of similar ability can be paired, and at other times weak students can be paired with strong students. The reader-writer pairings can have quite different purposes: from quick, personal responses to first drafts to editing and copyreading just prior to publishing in the class magazine or sending off to the school newspaper.

4. *In small groups of five or six students form a permanent writing workshop to give each other feedback.*

Many writers have shown how to use the writing workshop in English classes at both the school and college level (Elbow, 1973; Macrorie, 1970; Moffett, 1968). Some build the entire in-class writing program around it, while others use it only occasionally for particular purposes. With younger school-age kids it takes patience and persistence to get it started, but the effort does seem justified to me, particularly as class sizes grow larger. Since we can't get around to students often enough for writing conferences, I don't see any other way to provide the amount of formative response and feedback writers must have unless we perfect the writing workshop (and student pairs) as well as we are able in our classes. Two things are crucial: (1) training students to work effectively in small groups and (2) showing students what to look for in another's writing and how to respond helpfully. Two excellent sources of group training activities are Sax and Hollander's *Reality Games* (1972) and Stanford and Stanford's *Learning Discussion Skills through Games*. The ideas from Elbow and Macrorie above begin to illustrate what is involved in showing students what to look for.

After we have some acquaintance with the students and with their writing, we can form permanent workshop groups of five or six students. Each group can have the full range of writing abilities in the class. We can set out to get them acquainted and to train them to work together effectively. As often as we can, we sit in with each group and join the reading and discussing. In doing so we model for them what can be said about a piece of writing and what it means to be a good listener and a focused, helpful responder. We also provide this modelling when we lead whole-class discussion of student writing.

At first students can just pass around the circle the pieces they have written, or they can read aloud to each other what they have written. They can soon be given specific tasks: things to look for, proofreading, editing, *pointing, summarizing*. They can bring to the group only what they are willing to make public.

Some teachers find that the friendship or same-sex group works better than the teacher-arranged group, especially in Grades 7 and 8, where students are still aversive to the opposite sex, or at least acutely self-conscious with them. Some teachers also use same-ability groups.

However we arrange it and get it going the great value of the writing workshop is *quantity* and *immediacy* of feedback. The quality of the feedback is not very high, at least at first, and it is never as good as you could provide. But it does provide a ready and responsive audience besides the teacher, and it permits the writer to get almost immediate feedback the same day he writes or the next day.

Conclusion

Measurement and evaluation for writing involve finding out what the students can do when the course begins (diagnosis), responding to their efforts and guiding their growth (formative evaluation), and finding out how far they have come by the end of the course (summative measurement). In this article I have shown how we must redefine the conventional notions of mastery learning and formative evaluation to suit ourselves as English teachers, and I have outlined four general approaches to providing the continuous formative response and feedback young writers must have in order to grow as writers. There is nothing technical, complicated, or surprising in my recommendations. Our own good sense as teachers and writers already points us in the directions I have outlined. We know that composing is a complex personal, linguistic, and social process requiring much practice and supportive, helpful response. We know too, that growth is very slow and that we must be patient. With persistence and ingenuity we can overcome some of the school-as-institution constraints that frustrate us, and we can guide students' growth in writing as we know it must be done.

Most important, we must give up being the judge and become the coach or the listener. Sorting, grading, ranking, judging all subvert the writing process. We must concentrate on describing what students are doing and on encouraging their growth, and we must put off assigning the public grades we are still required to give until grade report time arrives. Even then, the grade should be based on willingness, on effort, and on growth.

Bibliography

Cooper, Charles R. "Measuring Growth in Writing," *English Journal, 64* (March 1975).

Elbow, Peter. *Writing Without Teachers.* New York: Oxford Press, 1973.

Judy, Stephen N. "Writing for the Here and Now: An Approach to Assessing Student Writing," *English Journal, 62* (January 1973), 69-79.

Klein, James. "Self-Composition," *College English, 35* (February 1974), 584-588.

Macrorie, Ken. *Telling Writing.* New York: Hayden, 1970.

McCallister, Lois, "Tell Me What You Had in Mind," *English Journal, 59* (February 1970), 231-234.

Moffett, James. *A Student-Centered Language Arts Curriculum, Grades K-13: A Handbook for Teachers.* Boston: Houghton-Mifflin, 1968.

Murray, Donald M. "What Can You Say Besides AWK?" *California English Journal, 9* (December 1973), 22-28.

Odell, Lee. "Measuring the Effect of Instruction in Pre-Writing," *Research in the Teaching of English, 8* (Summer 1974), 228-240.

Odell, Lee. "Responding to Student Writing," *College Composition and Communication, 24* (December 1973), 394-400.

Sax, Saville and Hollander, Sandra. *Reality Games.* New York: Macmillan, 1972.

Stanford, Gene and Stanford, Barbara. *Learning Discussion Skills Through Games.* New York: Citation Press, 1969.

Stanford, Gene. "Taming Restless Cats: Alternatives to the Whip," *English Journal, 62* (November 1973), 1127-1132.

SUNY, Buffalo

Recent Sources For Measuring Growth In Writing And Reading Fiction

Thomas G. Mac Lennan

While many of the books and articles listed below are fairly recent in origin, the concept of measuring growth in writing and reading fiction has been with us for some time. I'm sure that many teachers have puzzled over the problem of how to go about determining the status of a student who comes to you in September and leaves in June. Just what measures can we count on to determine the vital question of determining student growth from the beginning to the end of the period of time in question?

Measuring growth should not be confused with *evaluation*. The procedure of using growth measures to assign some number or letter grade which ostensibly reflects the quality of a student's performance is *evaluation*. *Measuring growth* concerns itself with the *valid* and *reliable* measures we use in collecting information about our students. If a measure is *valid*, it measures what it is supposed to measure. If it is *reliable*, it is fair and consistent with itself in measuring whatever it does measure.

There are a number of valid and reliable measures for writing and reading fiction and they are readily accessible. If you are located anywhere near a university or college library, they are available for your use as models in constructing your own growth measures. In some cases, I have added addresses just in case you want to add a copy to your own professional library.

MEASURING GROWTH IN WRITING

Adler, Richard. *An Investigation of the Factors Which Affect the Quality of Essays by Advanced Placement Students*. Ann Arbor, Michigan: University Microfilms, 1972. Can be ordered from University Microfilms, P.O. Box 1764, Ann Arbor, Michigan 48106. Order No. 72-12056.

Contains a very useful analytic scale for measuring growth in expository writing. I am going to be using the term, *analytic scale*, several times in this article, and Adler's is an example of this type of scale. The complete published scale describes the high, middle and low points for each component.

Quality of ideas	0	1	2	3	4	5
Development of ideas	0	1	2	3	4	5
Organization, relevance, movement	0	1	2	3	4	5
Style, flavor, individuality	0	1	2	3	4	5
Wording (choice of words)	0	1	2	3	4	5

Christiansen, Francis. "The Problem of Defining a Mature Style." *English Journal*. 57 (April, 1968), pp. 572-579.

An article that offers some suggestions in measuring growth in writing style. Christiansen suggests that the most valid measure of maturity is not sentence length, but length of the T-unit, which he defines as being each main or independent clause, together with its subordinate elements. He offers two hypotheses to be tested: (1) A mature style will have a relatively high frequency of free modifiers, especially in the final position. (2) Such a style will also have a relatively high frequency of structures of coordination within the T-unit. He differs with Kellogg Hunt's findings on syntactic maturity. Christiansen feels that a mature style must be easy to decode. The long clause isn't the mark of a mature style but of an inept style. The real problem is to pack much into little, "but to pack it so that it can be readily unpacked." For another view, see the Kellogg Hunt entry below.

Cohen, Arthur M. "Assessing Student's Ability to Write Compositions." *Research in the Teaching of English* 7 (Winter, 1973), pp. 356-371.

Contains an example of a dichotomous scale to be used in measuring growth in expository composition. This scale is useful when judgment doesn't have to be refined. An example, taken from Cohen's SCORE SHEET, follows:

	YES	NO	
Mechanics	____	____	There are many misspellings.
	____	____	There are serious punctuation errors.
	____	____	There are errors in the use of verbs.
	____	____	There are errors in the use of modifiers.

Cooper, Charles R. "Measuring Growth in Writing." *English Journal 64* (March, 1975).

If you only read *one* article about measuring growth this year, I would strongly recommend this one. It should be read for a number of reasons: (1) It offers the rationale for attempting to measure long term growth in a composition program; (2) it outlines some comprehensive objectives (including attitudinal ones) and a choice of *valid* and *reliable* measures you may select to attain the objectives; (3) implementation of the measures requires no special training,

and, with the exception of the attitude measures, are taken from a student's ongoing work and *do not* require separate tests or quizzes; (4) the objectives do not determine instruction; the classroom teacher is free to approach them anyway they see fit.
NOTE: The emphasis in this entire issue of the *English Journal* is on measurement and evaluation.

Corbin, Richard. *The Teaching of Writing in Our Schools*. New York: The Mac Millan Company, 1966.

You might want to examine NCTE's, "Ideaform" on page 84. It suggests an analytic scale (good, fair and poor) in measuring growth in expository composition. It is based on the following components: organization, development, sincerity of purpose and expression, sentence structure and punctuation, paragraphing, usage, choice of words, spelling, penmanship and general appearance.

Diederich, Paul B. "How to Measure Growth in Writing Ability." *English Journal* 55 (April, 1966), pp. 435-449.

This article not only concerns itself with *why* we should be measuring growth in writing but contains a model we can use in measuring growth of the whole school composition program. Diederich suggests using an analytic scale (high, middle and low) in measuring growth in General Merit factors (ideas, organization, wording and flavor) and Mechanics factors (usage, sentence structure, punctuation, spelling and handwriting). He strongly advises total department involvement in rating papers.

Diederich, Paul B. *Measuring Growth in English*. Urbana: National Council of Teachers of English, 1974. NOTE: Available from NCTE, 1111 Kenyon Road, Urbana, Illinois. Price is $2.50; member price is $2.25.

A most informative and valuable book. Diederich suggests several approaches to measuring growth in English: the reading test; the objective writing test; a short test on knowledge of grammar (the "shadow" test); as well as written essays which are to be graded by the entire department. Appendix A contains his analytic scale, described in the previous entry. By all means, read, "Initiating Staff Grading of Test Essays," on page 48, where he outlines the rationale and strategy for whole department grading.
NOTE: See Kenneth Watson's paper in this issue for a way to implement Diederich's departmental grading scheme.

Foley, Joseph J. "Evaluation of Learning in Writing." *Handbook of Formative and Summative Evaluation of Student Learning*. Edited by Benjamin A. Bloom and others. New York: Mc Graw-Hill Book Company, 1971. pp. 765-813.

Foley suggests that (1) we analyze the learning task; (2) organize the task into sequential units; (3) devise a table with desired cognitive and affective behaviors at the top and content (he lists ideas, organization, style, mechanics and choice of words) items on the side; (4) and then determine objectives by using the table intersection as a guide. He suggests a tripartite approach to writing: If you want to write, then read/write/think. By using the intersections of the table as a guide, the teacher can construct questions to measure growth. This essay gives numerous examples of how to develop measurement questions and contains a comprehensive bibliography.

Glazer, Joan I. *The Development of the Glazer Narrative Composition Scale*. Ann Arbor, Michigan: University Microfilms, 1972. Can be ordered from University Microfilms, P.O. Box 1764, Ann Arbor, Michigan 48106. Order No. 72-4496.

Contains an analytic scale designed to measure growth in student-written fiction. The scale has the following levels: plot, theme, setting, characterization and style. Each of these levels have several sublevels. For details refer to the measure. It utilizes many examples of student writing to illustrate the high, middle and low points on each level in the scale.

Hunt, Kellogg W. *Grammatical Structures Written at Three Grade Levels*. National Council of Teachers of English Research Report No. 3. Champaign, Illinois, 1965.

This research report deals with an instrument to measure syntactic maturity. The instrument was designed to study differences in sentence structures among a heterogeneous grouping of students of average I. Q. in the 4th, 8th and 12th grades. The degree of syntactic maturity was computed by counting the number of T-units (a T-unit is a clause and all its modifying phrases and subordinate clauses). The study determined that, upon graduation from 12th grade, a student writes T-units nearly 60 percent longer than he could eight years earlier. I strongly recommend Chapter 14, "Implications for the Curriculum," pp. 155-58. See also the Christiansen entry above for complementary measures of written syntactic fluency.

Judine, Sister M., ed. *A Guide for Evaluating Student Composition: Readings and Suggestions for the Teacher of English in the Junior and Senior High School*. Urbana, Illinois: National Council of Teachers of English, 1965. Available from NCTE, 1111 Kenyon Road, Urbana, Illinois 61801.

This collection of essays contains two composition scales. The California Essay Scale, pp. 147-158, concentrates on applying standards to content, organization and style. While it can be used for evaluation, the authors feel its best use is as a guide, "a reminder of the qualities which distinguish superior composition." The Cleveland Heights-University Heights (Ohio) City School District's, Composition Rating Scale, on pp. 159-160, is a five point, analytic scale based on content, style and conventions. It lets a student *see* what elements of his work need strengthening as well as the areas he has improved upon. Although it can be used as an evaluation device, it can also be used as a growth measure to show students where they are making progress.

Klein, Julie Thompson. "The Small-Group Approach to Writing." *Measure for Measure.* Edited by Allen Berger and Blanche Hope Smith. Urbana: National Council of Teachers of English, 1972. pp. 32-35.

This article suggests using student evaluation as an effective growth measure. Whole-class meetings are reserved for grammatical and stylistic discussions of model essays. By scheduling group conferences in a college composition course, the author found that the students gave each other a much broader audience than the student-teacher duo allowed. There was a significant indication that students were preparing and refining rough drafts of their compositions. This type of meeting also enabled students to exercise critical powers which could be then applied to their own work.

***National Assessment of Educational Progress—Writing Objectives.* Ann Arbor, Michigan: Committee on Assessing the Progress of Education, 1969.**

National Assessment is a plan for, "a systematic, census-like survey of knowledges, skills, understandings, and attitudes designed to sample four age levels (Age 9, 13, 17 and Adult) in ten different subject areas." Writing is one of the subject areas NAEP is concerned with, and writing exercises were developed to assess attainments specified by one of the following objectives:

(1) Write to communicate adequately in a social situation.
(2) Write to communicate adequately in a business or vocational situation.
(3) Write to communicate adequately in a scholastic situation.
(4) Appreciate the value of writing.

It would be very beneficial for anyone interested in measuring growth in writing to become familiar with the *objectives* and the *results* of the writing assessment. National Assessment has been, and to some extent still is, a controversial program. However, it does represent an attempt to determine, on a nation-wide basis, present status, past performance, and future direction. The following reports are available, for sale, from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402:

ORDER NO.	TITLE OF REPORT
3	Writing: National Results
5	Writing: Group Results A
8	Writing: Writing Mechanics
10	Writing: Selected Essays
11	Writing: Group Results A & B

NOTE: For two interesting views on the National Assessment of Writing Skills see Henry B. Slotnick's, "On the Teaching of Writing: Some Implications from National Assessment," and John C. Maxwell's, "National Assessment of Writing: Useless and Uninteresting?" in the December, 1973 issue of the *English Journal.*

Stoner, Don and Anderson, Art. "A Method for Teaching Subskills in Composition." *English Journal 58* (February, 1969), pp. 252-256.

An article that proposes an error-counting scale to measure growth in the compositional subskill of mechanics. The authors point out that the factors contributing to poor writing are:

(1) Prevelance of "busy work"
(2) Delayed return or non-return of student compositions
(3) Scales that measure composition as a whole

They suggest two possible models for measuring growth in the elimination of mechanical errors from expository compositions:

(1) For writing assignments of no more than a paragraph in length, they suggest a dichotomous scale in which a paragraph with no mechanical errors receives four and one or more errors receives a one.
(2) For writing assignments longer than a paragraph's length, they suggest a four point analytic scale (error free papers—4, one mechanical error—3, two mechanical errors—2, and three or more errors—1).

Tway, Eileen. *Literary Rating Scale*. Ann Arbor, Michigan: University Microfilms, 1971. Can be ordered from University Microfilms, P.O. Box 1764, Ann Arbor, Michigan 48106. Order No. 71-10997.

An interesting literary rating scale designed for use in rating elementary school (grades 4-6) student-written short fiction. It is an analytic scale which suggests high, middle and low points in the following elements: structure, ending, sentence structure, word usage, characterization, setting, point of view, conversation, detail, appeals to senses and values and situation.

MEASURING GROWTH IN READING FICTION

Cooper, Charles R. *Measuring Growth in Appreciation of Literature*. Newark, Delaware: International Reading Association, 1972. Available from International Reading Association, 6 Tyre Avenue, Newark, Delaware 19711. Members' price is $1.00, nonmembers' price $1.50.

A monograph written primarily for the researcher, but valuable to anyone interested in measuring appreciation of literature. A number of attempts to measure appreciation of literature are reviewed. Also contains an evaluation of limitations and possibilities of the measures and specific recommendations for further research in addition to a comprehensive bibliography.

Cooper, Charles R. and Purves, Alan C. *A Guide to Evaluation*. Lexington, Mass.: Ginn and Company, 1973.

More than just a complement to *Responding: Ginn Interrelated Sequences in Literature* this *Guide* contains several very comprehensive essays about measuring growth in response to a literary work. In addition, it gives teachers some valid and reliable measures for assessing growth. Cooper's essay, "Attitude Scales," explains the rationale and gives models of the Attitude-Sort and the Attitude Scale (pp. 22-27). Purves', "Observation of Responses," explains three systematic evaluation models. (1) For determining both the variety of responses and degree of involvement in those responses (model is shown on page 29). (2) A model based on observation of students' activities by adapting a behavior-content grid as an evaluation model (p. 30). (3) A classroom interaction chart (p. 32) which aids in assessing the interpersonal relations in the class.

Examples of the following measures are also included:

(1) Interests Questionnaire (pp. 37-38): Aids in the determination of students' likes/dislikes and interests.
(2) Preconception Measure (pp. 38-39): A questionnaire that deals with preconceptions students hold about what they read and what they see on television or in films.
(3) Reading Interests Questionnaire (p. 40): This questionnaire seeks to identify students' reading interests.
(4) Involvement Measure (pp. 41-42): Has the dual purpose of finding out (a) how readily students become involved in what they read and (b) what capacities they have for reading various kinds of literary works.
(5) Critical Approach Measure (pp. 42-43): Seeks to identify the characteristic ways in which students approach the reading of poems, plays, stories and essays.
(6) Student Profile (p. 45): A form that is used to synthesize the information obtained in the previous measures.

(7) Class Climate Inventories (pp. 47-52): Measures that provide classroom teachers a way to assess students' perceptions of your classes. For a discussion of these measures, refer to the Fox and others and Gorman entries in MEASURING GROWTH IN CLASS CLIMATE, below.

Gottesman, Julia M. "Making and Evaluating Literary Scrapbooks." *Measure for Measure*. Edited by Allen Berger and Blanche Hope Smith. Urbana: National Council of Teachers of English, 1972. pp. 54-55.

The author, a junior high school teacher of English, proposes an outline for making and evaluating literary scrapbooks. Her scrapbook evaluation contains a dichotomous scale to measure the mechanics, and an analytical scale to measure the content.

Loban, Walter; Ryan, Margaret; and Squire, James R. *Teaching Language and Literature*. New York: Harcourt, Brace and World, Inc., 1969.

This general methods text has a plan for evaluating growth in four important areas: ability to interpret behavior; sensitivity to form and style; grasp of ideas and theme; and growth of personal tastes. The section of the text devoted to evaluation of literature (pp. 562-569) contain some helpful models for constructing the following opinion polls: open-ended responses to a reader's response to a literary work, summary of voluntary reading, and self-evaluation instruments. There is a particularly good interest survey (p. 547) for gauging prevailing attitudes and tastes of students.

Moffett, James. *A Student-Centered Language Arts Curriculum, Grades K-13: A Handbook for Teachers*. New York: Houghton Mifflin Company, 1968.

Moffett feels that the administrative need for tests and marks, "has tampered with educational processes, especially in English, even more than most of its critics have ever asserted." In his "General Orientation" section of this volume, he does offer a method by which teachers can evaluate growth (p. 7). He feels that a folder of each student's papers should be kept and passed each year from teacher to teacher. Growth is measured by general assessment and teacher observation of the student's oral and dramatic work.

National Assessment of Educational Progress—Literature Objectives. Ann Arbor, Michigan: Committee on Assessing the Progress of Education, 1970.

For an account of the scope of the NAEP and where you may send for reports, see entry in MEASURING GROWTH IN WRITING above. The NAEP literature objectives are:

(1) Read literature of excellence.

(2) Become engaged in, find meanings in, and evaluate a work of literature.

(3) Develop a continuing interest and participation in literature and the literary experience.

In addition to a detailed explanation of these objectives and how they apply to the four age levels (9, 13, 17 and Adult), the report contains specific measures for literature. Anyone interested in measuring growth in literature should know the following reports:

ORDER NO.	**TITLE OF REPORT**
02-L-01	*Literature: Understanding Imaginative Literature* *(Theme 1)*
02-L-20	*Literature: Released Exercises*
02-L-02	*Literature: Responding to Literature* (Theme 2)
02-L-04	*Literature: A Survey of Reading Habits* (Theme 4)

Purves, Alan C. (ed.) *How Porcupines Make Love: Notes on a Response-Centered Curriculum*. Lexington, Massachusetts: Xerox College Publishing, 1972.

I highly recommend, as a good starting point in measuring growth, Purves' essay, "Proving That You've Done What You Set Out To Do And That It's All Been Worth It" (pp. 181-199). There is also a very comprehensive evaluation and planning bibliography on p. 216.

Purves, Alan C. "Evaluation of Learning in Literature." *Handbook of Formative and Summative Evaluation of Student Learning*. Edited by Benjamin A. Bloom and others. New York: Mc-Graw Hill Book Company, 1971. pp. 698-766.

In attempting to deal with the complex problem of constructing objectives, Purves suggests creating a grid with, "content on one axis and behaviors on the other, and determine from the grid those cells (or content-behavior intersections) which seem to be emphasized in a particular class, school, or curriculum." He includes a table (p. 701) to illustrate this process. The essay deals with detailed explanations of how to go about measuring response to literature. Purves shows how to construct measures of *knowledge*, usually questions of recall and recognition, which may be easily cast as multiple choice questions. He points out that questions of *application* take two forms, objective and essay, both of which ask the student to match two pieces of knowledge and determine significant relationships. He outlines questions of *expressed response* which look at the re-creative or the expressed response to pieces of fiction. Questions of *participation* seek to determine the relative weight of an interest in literature among a diversity of competing interests, and which seek to determine the depth of interest. The essay includes numerous examples of all of these measures.

MEASURING GROWTH IN CLASS CLIMATE

Doyno, Victor. "English Course Evaluation." *Measure for Measure*. Edited by Allen Berger and Blanche Hope Smith. Urbana: National Council of Teachers of English, 1972. pp. 97-98.

The author asks students for an anonymous evaluation of his course after about 5 weeks of class meeting. The evaluation centers around questions such as: What is your attitude toward the course? What should get more emphasis or attention? What could be done to improve the course? At the end of the course, he poses further questions to the class such as: What should be changed? What should remain pretty much as it was this semester? These final evaluations are also anonymous. Formative responses (*during* the course) are an aid to the instructor in case he wants to alter the emphasis of the course. Summative (*final*) responses aid him in teaching the course again.

Fox, Robert and others. *Diagnosing Classroom Learning Environments*. Chicago: Science Research Associates, 1966.

Includes over twenty five measures of classroom climate. Also includes the rationale for paying attention to students' responses of how they perceive classroom climate. This volume contains a questionnaire to determine how the class would like the teacher to act in the class (MY TEACHER). It asks students to give you some feedback on how they perceive you. Another measure, HOW THIS CLASS FEELS, provides information on students' perceptions about the class in general. PUPIL PERCEPTIONS OF A CLASS PERIOD provides information on students' perceptions of a single class meeting. NOTE: These measures are also contained in the Cooper/Purves, *Guide* above.

Gorman, Alfred. *Teachers and Learners; The Interactive Process of Education*. Boston: Allyn and Bacon, 1969.

A very helpful volume of measures of class climate. Contains a measure to gauge students' response to a single class meeting. Also provides the student the opportunity to comment on what he/she can do to improve the class (CLASS MEETING REACTION). There is also a very good measure of determining information on students' perceptions about the class in general (REACTIONNAIRE) which asks for student comments on the following questions: How would you describe your personal activity? What has been your reaction to the class thus far? Is it meeting your needs? Why or why not? How could this class be improved? What is on your mind that you wish you could say in class? NOTE: These measures are also contained in the Cooper/Purves *Guide* above.

SUNY, Buffalo

R + NR = ?

by
Patricia E. Moore
and
John T. Chamard

One of the eternal arguments in education deals with the validity of heterogeneous grouping. Our school has been involved in homogeneous grouping for many years. Four years ago we introduced an elective program in English for seniors. Some of us thought that these elective classes should be heterogeneously grouped much to the chagrin of several of our colleagues. These teachers thought that the better students would be held back with such an arrangement and that the less talented students would be inhibited by the presence of the more intellectual. We fought this argument because we did not think this would happen and furthermore these seniors would soon be going out into the world where it is necessary to communicate and work with people with a wide variety of educational abilities and backgrounds. We were certain that it was important that we do all that was possible to better prepare these students for the "outside" and that this was not being accomplished with classes designated A, regents, and non-regents.

We were able to influence the members of the English department enough so that they were willing to try it "our way." The transfer from homogeneous groups to heterogeneous was not easy for some students.

"We're just 'NRs' (non-regents). We can't do stuff like that," a senior from the back row groaned about the requirements for one ten week elective. Other voices joined in angry agreement. This was the first time in three years or more that our English students had been in mixed ability group classes. Later in an afternoon class, a very nice, very serious regents student came up after class and said, "I guess we won't do very much in this elective with *those* kids in here."

These feelings, however deep, came to the surface on and off through the year. There is no way to know if any changed their feelings about themselves or others, but it seems that three years of separation into A, regents and non-regents groups had an effect.

Tracking has been justified as the champion for "letting each become all he is capable of being." No more teaching for the "middle of the road," the high ability students are not held back, their intellectual curiosity is challenged by a curriculum that demands their best effort. The lower ability students are not frustrated by a curriculum that flaunts failure. They are not constantly reminded by the presence of brighter students that their answers are less substantial.

These are some of the positive points for tracking. (I remember the frustration of teaching a class of 30 ninth graders whose IQs ranged from 90 to 180.) The present system of tracking however, presents almost as many negatives. Lower track (NR) comes to mean: "I am not bright therefore I am not a valuable person. My ideas are not worth as much as those in the upper groups. I am not worth as much. I'm not going to try. What's the sense? Who will respect me for my accomplishments?"

Tracking is supposed to be based on ability, however, it is interesting to note that a majority of our lower track students have some common ground not shared so noticeably by the upper tracks: they are most often the students who break school rules; they often end up on detention, suspended, expelled or they quit school; they are seldom involved in extracurricular school activities; they don't usually hold offices in school organizations; their over-all attendance is poor; and, in general, they lose most popularity type contests. There are also a number of students who end up in non-regents classes who are not incapable of doing regents work, but because of laziness or discipline problems they have wangled their way into a class that is easy for them. If they were again forced into a class with the better students, might they not benefit with more acceptable work and discipline due to peer pressure?

These observations and the responses to our heterogeneously grouped senior electives prompted us to consider mixing the sophomores for two short units during the year. We outlined a plan for a research study to use four of our classes in the spring for two, two week units. The first was a speech unit and the second a travel unit requiring reading and writing skills. We combined two classes (one regents, one non-regents) into the experimental group, which was divided into two sections. The other two (also one regents, and one non-regents) remained intact as our control groups.

We drew names out of two hats to create the two experimental groups. Each group had approximately thirteen regents and thirteen non-regents students. We gave the students a simple explanation about being tired of the same groups and sent them off to the other class. No one questioned us. A few grumbled about having to get up from their comfortable slouches to go to a new room and a few others looked a little anxious, however, they did cooperate.

The first unit required each student to prepare and present to the class a demonstration speech on a topic of his/her own choice. It is our feeling that this type of speech is least frightening since the speaker has something to do with his hands, a specified process to follow, and an end result to show. All students had done other types of speeches earlier in the year. Topics chosen in this unit included changing a flat tire, making Jello, tie-dying T-shirts, fly fishing methods, candlemaking and a variety of sports demonstrations.

All students were rated on a 1-5 scale in four areas: opening, development of point, conclusions and delivery. The results showed no significant difference between the control and experimental groups in any of the areas or in the totals. The evidence of the control group seems to say that the regents students in the experimental group functioned at a normal level; they didn't try less because they received little or no competition, but rather they worked at their usual level. This was also true of the non-regents. The experimental group was not inhibited by the regents students and continued to work at their usual level.

In general, students in both groups feel pretty much the same about speeches, they don't like them. The non-regents students seem to hate them less than the regents, perhaps because oral communication is emphasized in their curriculum. The best score in one of the experimental groups was made by a boy regularly in the non-regents class. At the same time there were three non-regents boys in the experimental and control groups who stayed out of school most of those two weeks to avoid the speech; they had also done this earlier in the year before this experiment.

The second unit, the compiling and writing of a travel brochure, started right after our two week spring vacation. This unit required the students to set-up a $2000 vacation pamphlet in process form. Each student chose his destination, computed the distance and cost, decided on sights to see, found accommodations and costs for everything and organized a daily itinerary. Some students were quite creative, some of them were not. This was an ambitious project and the ambitious students in both groups were the more successful ones.

For the brochure, students were rated on a 1-5 basis in seven areas; attractiveness and usefulness of cover, general information, organization and detail of text, overall appearance, and originality. The first and sixth categories were meant specifically to give someone credit who tried hard but had difficulty with reading and writing skills. The statistical data showed no significant difference between the experimental and control groups in any of the areas.

Both of these units required individual research, preparation and performance. The first day in each unit consisted of general lecture, instruction from then on was individualized. Perhaps performance of the experimental and control groups was similar for this reason. Perhaps this in part accounts for the fact that academic success in these units did not directly correlate with track level. This is the way we hoped it would work out.

We were concerned that mixing the classes for the experimental group might cause the non-regents to feel less successful than in their original groups. That is why we introduced the CALIFORNIA TEST OF PERSONALITY, A PROFILE OF PERSONAL AND SOCIAL ADJUSTMENT which measured feelings in six areas; self-reliance, sense of personal worth, sense of personal freedom, feeling of belonging, withdrawing tendencies and nervous symptoms.

The day before the new class groupings were announced we gave a modified version of this self-image test. No names were used on these, instead students identified their tests with their birth dates. This same method was used on the post-test so that changes in responses could be noted. The pre-test scores for the regents students showed they already had a more positive self-image than the non-regents students. It was not surprising to find that the regents student felt better about himself than the NR. At the end of the experiment, when the students again took the California Personality test, the data showed some significant changes.

On variable five, withdrawing tendencies, the regents control group showed an improvement from the pre- to post-test that proved to be mathematically significant (7.14 to 8.00). Although no other individual variable proved significant for this group, they showed a measurable improvement of self-image at the end of the experiment (52.81 on pretest to 55.29 on the post-test).

In the area of nervous symptoms, variable 6, the regents experimental group showed a significant improvement (8.36 to 9.43). Again, even though the other variables showed no significant improvement, the test as a whole showed significant improvement of self-image (52.18 on pre-test to 55.04 on post-test).

The non-regents control group showed no significant differences in any of the individual areas or on the self-image test as a whole.

Although the non-regents experimental group on variable six, showed a significant difference (9.07 to 8.13) which indicates a greater tendency toward nervous symtoms, the overall test totals for this group showed no significant difference.

It appears that academically the two groups functioned in their usual manner. The inclusion of a higher ability group did not frighten or stifle the lower ability students while the regents students did not lower their standards. The individualized type task required in this unit kept class interaction at a minimum. Class discussion, small group work, formal testing were not a part of this unit. These types of tasks may be more threatening than those used in these units.

An overall look at the self-image appears to show regents students in both the experimental and control groups feeling better about themselves while the non-regents self view did not change significantly. We really did expect a change for the non-regents one way or the other. Maybe they have bottomed-out by the time they reach sophomore year, maybe only a continuous mixing will have any effect, maybe they've built such solid walls that only dramatic failure or success can effect them now, maybe once you're labeled, you're labeled forever. It undoubtedly would take more than a four week experiment to produce a significant improvement in self-image for these non-regents students. It must be stressed, however, there was no decrease in self-image because of their exposure to the better students which supports our theory that NRs will not feel inferior because of exposure to Rs.

We had some personal observations not supportable by data which were interesting to us. There were several troublesome students in the non-regents experimental class before we mixed them, who modified their behavior to something more teacher-acceptable. One non-regents boy in the experimental class who had been totally silent all year made friends with three girls from the regents class, ended up dating one and started participating in class discussions. One shy quiet young lady in the non-regents class became one of the most popular students in the experimental class. This observation by the teacher was further supported by a socio-gram given at the end of the experiment.

For this socio-gram students in the experimental classes chose five others from the class with whom they would like to work in the next unit of study. We discovered that all students were included in the socio-gram and there was a great deal of mixing of regents and non-regents students. This proved to us that there was much to be gained socially in mixing the two ability groups.

It seems there are gains to be made and little to be lost in breaking grouping barriers for short units of this type at this level. As a result of this experiment, we are now involved in a fifteen day unit that includes A, regents, and non-regents sophomore students. This grouping concept will be repeated at least once more during the present school year. Unfortunately it isn't yet a reality for the junior year. We can only guess about senior electives for these sophomores. Obviously, much will depend upon the kind of experiences they have in these short mixed units. Faces and names will be familiar which may help if the known is really less frightening than the unknown.

Hopefully arrival in the non-grouped senior elective program will not be such a shock for these students, and entrance to the real world of work and life may be easier for those who must cross all kinds of group barriers to succeed. Industry, government, education depend on all types of skills, abilities and personalities for their success. Management must be able to communicate with labor and vice versa. If we compartmentalize, departmentalize, track and, in fact, build walls in the education that must (should) be broken down outside these brick edifices than we are not educating whole people for whole living. The neighbors next door, are they A's or NR's does it matter?

Oneonta High School
Oneonta

Putting Cloze Into The Classroom

by John F. Haskell

ESL teachers regularly face the problem of selecting reading materials that will meet their student's needs and abilities. Materials prepared for ESL students are often designated as being at some word level such as the 1,000-word level or the 2,500-word level or are prepared for more general use with groups known as "intermediate" or "advanced" level students. But it is difficult to determine accurately where the students are in relation to those levels. Readability formulas, which might help, are not appropriate for classroom use as they are both too complicated and too time consuming to use. In addition, readability formulas are generally designed to measure materials used with native speakers and are interpreted in terms of grade level. School grade levels are not easily applied to ESL students and certainly not with the same meaning.

In intermediate or advanced reading classes it is often the practice to ask students to select something to read outside of class. This may seem to be a way of providing motivation for reading—allowing the student to choose for himself—but, how many times have each of us stood in front of a rack of books in the airport or at the drugstore and been unable to select a book to read—sexy cover or no?

What I wanted to find, in my attempt to solve this problem of selecting reading material, was an efficient method that any classroom teacher could use. A method to measure materials for level of readability which would also help the teacher determine the appropriateness of the materials for use with a specific group of students. It had to be also, easy to prepare, easy to give, and easy to score. What I found was the CLOZE procedure.

The Cloze procedure for measuring readability was first suggested by Wilson Taylor as a means of evaluating materials for native English speakers in 1953.[1] It has since been used with Japanese, Korean, Amharic and other languages as well as ESL. In looking at various studies of Cloze with ESL students, to determine which procedures would be the most efficient and still practical to use in the classroom, I found a wide variety of procedures used, very few of which were selected on the basis of any tested evidence.

The Cloze procedure is basically the deletion of every Nth (5th, 7th, 10th or the like) word from a portion of the material to be evaluated, which is then, in its mutilated form, given to the student. The student is asked to restore the text by replacing the blanks with the appropriate words—by guessing, from the context. The theory is that his ability to restore the test is a measure of the appropriateness of the material for the student to read. Except in the cases of blank length and type of scoring, I found no research reported to determine Cloze procedures to be used with ESL students.

As a result, two studies to evaluate Cloze testing and scoring procedures were undertaken to determine which would be most useful to the classroom teacher in selecting reading material for ESL students.[2] The following statements are based on conclusions drawn as a result of those studies.

1. Deletion Rate. Many researchers, including John Carroll and Dorothy Danielson seemed to feel that foreign students should be given as much information between the blanks of a Cloze test as is practical (they have used a deletion rate of every 10th word). I found, in testing various rates, that, as with native speakers, passages deleted of every 5th word produced the same statistical results, that is, provided the teacher/researcher with the same information about the readability of the passage, as when those passages were deleted of every 7th word or every 10th word. Students did not tend to score any better when they were given nine words between each blank than they did when they were given only four words between each blank. (Fewer than four words between blanks, however, did affect results.)

For classroom use, the deletion of every 5th word is, then, the most efficient or economical deletion rate because, for a standard fifty deletion passage, it allows the teacher to use a shorter passage and still maintain a sufficient number of deletions for measuring. The time needed for test preparation is also less when only every fifth word is deleted.

2. Deletion Type. The mechanical deletion of every Nth word is what makes the procedure work as a readability or proficiency testing instrument. Selective deletion (that is, deleting only nouns, prepositions, or the like) changes the character of the test from one which tests total language competence to one that tests the ability of the student to recognize and use words in a grammatical context.

Some researchers have been bothered by the fact that certain numerical figures and proper nouns, when deleted, were nearly impossible to guess correctly. With native speakers, when unguessable numbers or proper nouns turn up in the process of deletion, they are generally skipped and the next word deleted. Counting continues from the deletion. This procedure does not disturb the mechanical nature of the procedure, when used with ESL students, nor the value of the test results. But it was also found that the inclusion of two or three such unguessable words, on any given fifty deletion test, did not disturb the test results—especially when they are used to evaluate group scores.

3. Number of Deletions. Most of the studies with native speakers recommend a minimum of fifty deletions for a Cloze test. In one study with ESL students I used passages with a range of thirty-five to seventy-five deletions. Given the materials used (selections from ESL readers),

the number of deletions, within this range, did not seem to affect test results. However, test scores based on fifty deletions proved to be more easily converted to percentage scores. Fifty deletions also meets most research needs for length.

4. Test Length. Given the procedures recommended above (a deletion rate of every 5th word, mechanically deleted, with a minimum of fifty deletions) the average test length should be approximately 250 words. (Approximately, since it is always best to take the passage to the logical end of a sentence rather than abruptly end with the fiftieth deletion.) For the classroom teacher this means that a single test passage can be typed, double-spaced, on a single page.

5. Blank Size. The studies done on Cloze testing procedures confirmed Jonathan Anderson's previous research which stated that the length of the blank which replaced the deleted word did not affect the results of the test.[3] That is, blanks of either uniform length or various lengths (either conforming in some way to the length of the word deleted or to the requirements of the typewritten page) can be used. Furthermore the teacher may even blank out the words on the original material rather than type up the passage and either test it that way, or, for a group of students, have the page reproduced by some mechanical process.

6. Scoring Procedure. Research on Cloze scoring procedures with ESL students have dealt with three basic scoring methods; "Verbatim" (or exact-word) scoring, "Synonym" scoring, and "Other-word" scoring. In Vertabim scoring only those words that occur in the original test are acceptable. In Synonym scoring, synonyms of the exact words are also accepted. In Other-word scoring, any appropriate word that occurs and continues the general meaning of the original text is accepted. In all cases spelling is not considered important.

For the evaluation of scores for a group of students on any given piece of material, Verbatim scoring provides about the same information about the READABILITY of the passage as the other two scoring methods and is therefore preferred for use by the classroom teacher because it is more objective and takes less time to score. This conclusion has more recently been confirmed by Krashen, Zelinski, and Jones[4] and in a personal communication by John Oller[5].

We assume in testing, that objectivity is what we are aiming for. I found that for measuring the readability of a passage for a group of students, the Verbatim scoring method was the most objective and the most efficient to work with. (This is particularly true when the blanks on the deleted passage are numbered in some way.) There is no need to evaluate each answer for its appropriateness as is the case in both of the other scoring methods discussed here. A single list of the deleted words can be used instead.

On the other hand, for interpreting individual scores, objectivity may not be the important factor. Individual scores may vary widely due to the differences in knowledge of vocabulary, or grammar, experience or interest. That some individuals, for various reasons, may have more or less success with the material may not be as important to the teacher as the over-all appropriateness of the material for the whole class or group of students. But for individual diagnosis of a Cloze test the Other-word scoring method provides a greater amount of information with which to analyze the student's strengths and weaknesses. Other-word scoring is more subjective and often not scored consistently but it does provide an indication of vocabulary strength and versatility as well as helping the teacher determine specific grammatical and semantic problems of individual students.

7. Evaluation. No research has been reported that establishes evaluation criteria for Cloze readability testing with ESL students, but some useful generalizations can be made from procedures used for native speakers. Jonathan Anderson states, without providing any research information, that he found the general criterion levels established for native English speakers to be appropriate for ESL students.[6] In general they are the following: scores of 53% or above suggest that the material is suitable for the student to read independently, scores of between 44 and 53% indicate that the material is suitable for instructional use—in the classroom, and scores below 44% indicate that the material is too difficult for the student and will only produce frustration. These are general areas and the teacher may want to adjust them somewhat.

It should also be kept in mind that the newness of the Cloze procedure itself may affect scores the first time or two it is used. Most of us are used to the idea that 100% is possible and the aim of the student in testing. The Cloze passage is not a test in this sense and it should be made clear, that, in general, it is being used to measure the READABILITY of the material and not the reading ability of the student. A score of 100% is highly improbable, even for a native speaker, and it is the score of between 44 and 53% that the teacher is looking for in terms of the appropriateness of the materials for use in the classroom (or scores above 53% for independent reading).

The Cloze procedure provides an economical and efficient method by which the teacher can select materials for a class of students or for diagnosis or evaluating an individual students reading problems. It seems to me, that it could also be put into the hands of the student as a method by which he might learn to evaluate his own reading selections and perhaps even to evaluate his own reading progress.

In summary, there are simple, easy, and efficient procedures which the classroom teacher can use in preparing a Cloze passage to be used as a readability measure. (1) Select about 250 words from the material to be evaluated. (2) Delete mechanically every fifth word, beginning at the fifth word and skipping proper nouns and numbers should more than three occur. (3) Delete fifty words. (4) Replace the deleted words with numbered blanks. (5) Score the passage by accepting only the words from the original text (Verbatim scoring). (6) Evaluate scores in general categories of Frustration (scores below 44%), Instruction (between 44% and 53%), and Independent reading (above 53%).

It is hoped that the classroom teacher, understanding the generality of the evaluation criteria will, nevertheless, take advantage of this easy to prepare, easy to give, and easy to score readability measure and try it out. Get students to use it, get them used to it, and set up some evaluation criteria that will work for their particular class—but use it.

SAMPLE CLOZE TEST AND INSTRUCTIONS

INSTRUCTIONS: On the next page is a sample of a new kind of test. Some of the words are left out of the story and blanks are put where the words were taken out. YOUR JOB WILL BE TO GUESS WHAT WORD WAS LEFT OUT OF EACH BLANK and to write the word in the blank.

Note: The last part of this sentence might read 'and to write the word on a separate numbered piece of paper.'

The following sample sentences may be helpful for students who are using this type of reading passage for the first time. (The answers are written upside-down under the sample sentences.)

1. This ____ a book.
2. I ____ like ice cream.
3. I ____ ice cream, but I don't like candy.
4. I like ice cream, ____ not candy.
5. I like ____, but not ____

Note: These can be done by the students silently and then discussed orally. Ask students for different answers that might have occurred in numbers 2. and 5.

1. *is* 2. *like, hate, love,* etc. 3. *like* 4. *but* 5. almost any noun or pronoun.

Instructions for the next page:

1. Try to read the entire passage through once before you begin to WRITE in the blanks.
2. Write only ONE word in each blank. (Contractions are possible.)
3. You may skip hard blanks and return to them later.
4. Spelling will not be counted.
5. READ THE ENTIRE PASSAGE even though it may seem difficult. You will be able to fill in some blanks somewhere on the page.
6. Try to fill in each blank. DON'T BE AFRAID TO GUESS.
7. This is NOT a test. You will not be graded on what you do.

SAMPLE CLOZE PASSAGE: From "The Casting away of Mrs. Lecks and Mrs. Aleshine," by Frank R. Stockton and adapted for ESL students in *People in Fact and Fiction* by Virginia French Allen.

I was on my ____ (1) from San Francisco to Yokohama ____ (2), in a very gradual ____ (3) I became acquainted with ____ (4) Lecks and Mrs. Aleshine. ____ (5) ship on which I ____ (6) traveling at the time ____ (7) a large number of ____ (8), most of whom were ____ (9) Among these, I noticed ____ (10) the very first day ____ (11) middle-aged women who ____ (12) different from the ordinary ____ (13) or tourist. At first ____ (14) they might have looked ____ (15) farmer's wives who, for ____ (16) unusual reason, had decided ____ (17) cross the Pacific. On ____ (18) observation, however, it seemed ____ (19) probable that they belonged ____ (20) the families of prosperous ____ (21) in some little country ____ (22); there, in addition to ____ (23) and sewing, the women ____ (24) have opportunities of becoming ____ (25) with some of the ways and manners of the outside world.

(1) way (2) when (3) manner (4) Mrs. (5) The (6) was (7) carried (8) passengers (9) Americans (10) on ((11) two (12) were (13) traveler (14) sight (15) like (16) some (17) to (18) closer (19) more (20) to (21) shopkeepers (22) town (23) cooking (24) would (25) acquainted.

---

[1] Wilson L. Taylor, "Cloze Procedure: New Look for Measuring Readability, **Journalism Quarterly**, XXX (1953), 115-33.

[2] John F. Haskell, "Refining Cloze Testing and Scoring Procedures for Use With ESL Students" (dissertation, Teachers College, Columbia University, April 1973).

[3] Jonathan Anderson, "Techniques for Measuring Reading Comprehension and Readability," **English Language Teaching**, XXV (February, 1971), 178-81.

[4] Stephen D. Krashen, Stanley J. Zelinski, and Carl M. Jones, "Report on a Cloze Test," **Idiom** IV (Summer, 1971).

[5] John W. Oller, Jr. March 8, 1974.

[6] Jonathan Anderson, "Selecting a Suitable Reader", **RELC Journal**, II, December, 1971, 36-42.

A Bibliography of Articles and Papers Relevant To Cloze Testing In ESL and FL

Allen, Virginia French. "Towards a Thumbnail Test of English Competence." *TESOL Quarterly*, II (March, 1968), 241-46

Anderson, Jonathan. "Selecting a Suitable 'Reader': Procedures for Teachers to Assess Language Differences." *SELC Journal*, II (December, 1971), 35-42.

________. "Techniques for Measuring Reading Comprehension and Readability." *English Language Teaching*, XXV (February, 1971), 178-81.

Bosco, Joseph. "Readability and the Selection of Materials for Adult Learners: A Procedure." New York: Puerto Rican Forum, 1973 (mimeographed).

Bowen, J. Donald. "A Tentative Measure of the Relative Control of English and Amharic by Eleventh Grade Ethiopian Students." *Workpapers in TESL*. Volume II. Los Angeles: University of California, 1969.

Carroll, John B.; Carton, Aaron S.; and Wilds, Claudia P. *An Investigation of Cloze Items in the Measurement of Achievement in Foreign Language*. Cambridge, Massachusetts: Laboratory for Research in Instruction, Graduate School of Education, Harvard University, 1959.

Darnell, Donald K. "Clozentropy: A Procedure for Testing English Language Proficiency of Foreign Students." *Speech Monographs*, XXXVII (March, 1970), 36-46.

Freidman, Mildred. "The Use of the Cloze Procedure for Improving the Reading Comprehension of Foreign Students of the University of Florida." *Dissertation Abstracts*, XXVII (December, 1964), 3420.

Gregory-Panopoulos, John Fred. "An Experimental Application of Cloze Procedure as a Diagnostic Test of Listening Comprehension Among Foreign Students." *Dissertation Abstracts*, XXVII (January, 1967), 2231.

Grinsell, Kathleen. "What Test and Why." Paper delivered at the 1973 TESOL Convention in San Juan, Puerto Rico, May 11, 1973 (mimeographed).

Haskell, John F. "Refining the Cloze Procedure for English as a Second Language." *English Record*, XXV (Winter, 1973), 77-82.

________. "Read This: But How Do I Know They Can?" *TESOL-GRAM*, III (February, 1975), 2, 4.

________. "Refining Cloze Testing and Scoring Procedures for Use with ESL Students." (dissertation, Teachers College, Columbia University, April, 1973).

________. Selecting Reading Materials: A New Approach." *Idiom*, III (Fall, 1972), 1, 4.

Holtzman, Paul. *English Language Proficiency Testing and the Individual*. University Park, Pennsylvania: Pennsylvania State University, 1967.

Kaplan, Robert B. and Jones, R. A. "A Cloze Procedure Test of Listening for University Level Students of ESL." Los Angeles: University of Southern California, 1970. (mimeographed).

Krashen, Stephen D.; Zelinski, Stanley J.; and Jones, Carl M. "Report on a Cloze Test." *Idiom*, IV (Summer, 1974).

Klare, George R.; Sinaiko, R.W.; and Stolurow, L. M. *The Cloze Procedure: A Convenient Readability Test for Training Materials and Translations*. Arlington, Virginia: Institute of Defense Analysis, 1971.

Mason, Victor. "Report on Cloze Tests Administered to Thai Students." *Bulletin of the English Language Center*, II (March, 1972), 31-50.

Oller, John W., Jr. "Cloze Tests of Second Language Proficiency and What They Measure." *Language Learning*, XXIII (June, 1973), 105-18.

________. "Dictation as a Device for Testing Foreign Language Proficiency." *English Language Teaching*, XXV (April, 1971), 254-59.

________. "Level of Difficulty and Scoring Methods for Cloze Tests of English as a Second Language Proficiency." *Modern Language Journal*, LVI (March, 1972), 151-58.

________; Bowen, J. Donald; Ton, That Bien; and Mason, Victor W. "Cloze Tests in English, Thai, and Vietnamese; Native and Non-Native Performance." *Language Learning*, XXII (June, 1972), 1-6.

________, and Conrad, Christine A. "The Cloze Technique and English as a Second Language Proficiency." *Language Learning*, XXI (December, 1971), 183-95.

________, and Inal, Nevin. "A Cloze Test of English Prepositions." *TESOL Quarterly*, V (December, 1971), 315-26.

________, and Redding, E. Z. "Article Usage and Other Language Skills." *Language Learning*, XXI (June, 1971), 85-95.

Pack, Alice C. "Cloze Testing and Procedure." *TESL Reporter*, VI (Winter, 1973), 1-2.

Shiba, Sukeyori. "A Study of Readability Measures—Application of the Cloze Procedure to the Japanese Language." *Japanese Journal of Psychology*, (English Abstract), XX-VII (1957), 67-73.

Taylor, Wilson L. and Lee, K. W. "KM Readers Lend Hand to Science: 'Cloze' Method Works in Written Korean and May Serve as a Tool for Korean Language Reform." *Korean Messenger*, III (1954), 4-5.

Note: Various unreported studies are being carried on throughout the country. Among those which have come to my attention are at the State University of California: San Francisco (Prof. Dorothy Pilgrim Danielson) and The Learning Center, Jersey City State College (Dr. Phillip Shew).

University of Puerto Rico

Reading Frost:
"The Road Not Taken"

Roberts W. French

Is there any poem in American literature more often and more consistently misinterpreted than Robert Frost's "The Road Not Taken"?

Again and again the final three lines are used not only to describe Frost's life, but also, and more generally, to characterize anyone who has broken with convention and set out on a lonely, independent course of action. For students these concluding lines are particularly attractive, since they seem to be a succinct and lucid summary of the poem's "message"; and thus the poem appears to reinforce the attitude that many students bring to poetry, that poetry is nothing more than circuitous language perversely concealing plain statement.

A careful reading of the poem, however, will show that it is by no means the ringing affirmation of independence it is often taken to be, but, rather, a poem of defeat and failure. What's more, a careful reading will show that the standard interpretation can only be derived from a persistent refusal to see what the words are actually saying. Most students, like their elders, will ignore the poem in favor of their own preconceptions; and therefore this poem is particularly apt for classroom teaching, since it demonstrates clearly the faulty perceptions that arise when we neglect what might be called the first law in reading poetry: *look at the words.*

When we look at the words, we should note that the poem contradicts itself in a curious and significant way (a profitable class discussion might begin by asking students to explain this contradiction). The speaker tells us in the famous last lines that he took "the road less traveled by," thus implying that he followed an independent, adventurous way of life, perhaps at some risk or cost to himself. Earlier in the poem, however, the speaker tells us that the roads were approximately the same, with no essential difference between them. One was "just as fair" as the other, and as for travel, "the passing there/Had worn them really about the same,/And both that morning equally lay/In leaves no step had trodden black." Were the roads different, as in the last stanza, or much alike, as in stanzas two and three? Why can't the speaker make up his mind?

Well, he can, and has. Does he *in fact* say that he took the road less traveled by? It is generally assumed that he does, but the words of the poem say something rather different. The speaker does *not* say that he took the road less traveled by; he says that he *will say*, in the distant future ("Somewhere ages and ages hence"), that he took the road less traveled by. The first two lines of the last stanza make this distinction clear, but they are usually ignored as though they did not exist. The speaker is by no means looking back on his life and commenting with pride on his rugged independence. For one thing, he is neither independent nor decisive; as the first stanza tells us, he is a long time making up his mind, and even when he has done so, he likes to think that the choice is not final, that he may yet be able to return and choose again ("Oh, I kept the first for another day!"). Furthermore, the speaker can hardly be looking back on a life nearing its end, since his reference to "ages and ages hence" indicates that he still has much life to live. To say this, however, raises several questions: since the speaker's future still lies wide open before him, how can he know what he will be saying near the end of it? What makes him so ready to offer premature judgments on the whole course of his life? What "difference" is he talking about? And why will he say that the roads were dissimilar, when he has clearly revealed that they were practically identical?

Back to the text. The speaker notes that when he will say that he took the road less traveled by, he will say it "with a sigh." If he were asserting his individualism, he would certainly not be doing it "with a sigh." For what are the implications of a sigh? What mood does it suggest? Not happiness, surely, nor confidence, nor pride, but something like regret, or melancholy, or wistful sorrow. Contributing to this mood is the curious phrase, "Somewhere ages and ages hence." Why would anyone refer to his future in just these words? The phrase suggests, among other things, that the speaker sees his future as extremely long. We are accustomed to think of life as too short, in no way extending beyond us for "ages"; so that if one's future can be conceived in terms of "ages and ages," one must indeed be weary of living. In any case, the speaker can hardly be said to face the future with enthusiasm or eager anticipation; the future seems, rather, to be something of a burden.

What, then, are we to make of this person? We know that he has, of necessity, made a choice, one "road" instead of another, and that the choice will have significant bearing on the course of his life (the last stanza makes clear that he is talking about more than roads). Furthermore, we know that he is going to tell a lie: he is going to say (he has told us so) that he took the road less traveled by when in fact both roads were about the same, equally attractive and equally worn. Why is he going to lie? And how does he *know*, so soon, that "Somewhere ages and ages hence" he is going to lie?

Only reference to the details of the poem can answer these questions; but it is just these details that are often overlooked. When they are all before us, however, they form a consistent pattern. In the end, it becomes apparent that the speaker is preparing his excuses; he expects to fail, and having done so, he will blame his failure on the independent course of his life. By claiming to have taken the road less traveled by, he will be implying that he has chosen to avoid the world's ways and therefore could not be expected to succeed on the world's terms. He will be lying, of course, for we know that he did *not* take the road less traveled by; but the lie will provide a respectable excuse for the failure that he sees as certain and inevitable—indeed, if the lie is believed, it may even make failure look like heroism, which is no doubt the speaker's intent. He is anticipating a life of failure and defeat, with a sigh at the end of it; no wonder the future seems so long.

As this brief reading suggests, Frost's poem is more intricate and more complex than the popular understanding of it would indicate; it is also better: more subtle, more perceptive, more analytical, more deeply concerned with human motivation. As a subject for classroom discussion, "The Road Not Taken" has many virtues, not the least of which is its apparent simplicity: it seems so easy. The poem is deceptive, however, for only the language is simple, not the technique. Still, its accessibility encourages exploration; and as it is explored, the poem reveals itself, piece by piece, until the pieces form a coherent whole. It is not, then, a *difficult* poem so much as a *demanding* poem: in order to read it at all, we must sit up and pay the closest attention to detail, if the poem is not to be lost. It is a poem that can teach us, finally, that most important of lessons, the necessity of taking language seriously.

University of Massachusetts

A Development Cycle for a Competency Based English Curriculum, Grades K-12

by Lester S. Golub

The rapid economic, social, and political changes of the decade of the 70's have placed strains and stresses on the "traditional" elementary and secondary school English curriculum and English teachers. In an age which demands advanced literacy skills for survival of even the truck driver and the fireman, English teachers in the schools are *accountable* to more than the principal of their school and the district superintendent. Teachers of English are responsible to the American society which hires them to do a job. This present American society has expectations for the upcoming generation. Responsible adult members of this society expect the entering members of the society to be *competent* in carrying out certain carefully delineated *behaviors* and *thoughts* which satisfy the *needs* of the society. In developing an English curriculum, K-12, for the children and adolescents of an American school district, teachers must consider: (1) their accountability to the society which hires them and pays their salaries; (2) the needs and expectations of that society; (3) the carefully stated English and reading objectives determined to meet these needs, and (4) the attained and developing English and reading competencies of the children and students as they progress from level to level in an instructional system.

In the development of an instructional system in English and reading, elementary and secondary, the children and students are perceived as developing humans with an expanding repetoire of cognitive and behavioral abilities and experiences. This stance is the first break from tradition since developmentally, children and students progress from level to level of competency, rather than from grade to grade. Levels, however, encompass grades as follows:

Level I,	Grades K-2,	Ages 5-7	Ego-centric cognitive and behavioral ability
Level II,	Grades 3-4,	Ages 8-9	
Level III,	Grades 5-6,	Ages 10-11	Concrete cognitive and behavioral ability
Level IV,	Grades 7-8,	Ages 12-13	
Level V,	Grades 9-10,	Ages 14-15	Abstract cognitive and behavioral ability
Level VI,	Grades 11-12	Ages 16-17	

At each level, the content areas include: (1) listening, (2) speaking, (3) reading, (4) writing, (5) language, (6) literature, and (7) media. Certainly, not all of the curriculum fit under the format of a competency based English curriculum, but an attempt at a competency based English curriculum at each level leads to better understanding of ways of assuring students' competence in English and reading performance. The developmental stages (Figure 1) to be encountered over a period of two or more years include: (1) needs assessment; (2) selection and description of concepts and skills to be developed at each level; (3) development of concept and skill objectives; (4) development of test items for concept and skill objectives; (5) development of materials and activities for teaching objectives; (6) development of individualized, learning activity packets which include (a) objectives, (b) pre-test, (c) learning activities for each objective, and (d) post-test; (7) formative evaluation and revision of Learning Activity Packets for maximum student attainment of objectives; (8) development of the sequence and scope of a Competency Based English Curriculum at Levels I, II, III, IV, V, and VI. Each of these developmental stages requires careful explanation and deliberation.

Needs Assessment—The needs assessment is the first and most critical stage in the development of a competency based English curriculum, K-12. The needs assessment relies on input information from the following sources: (1) knowledge of students' future needs in school, society, and culture; (2) knowledge available from the psychology and sociology of learning; (3) knowledge from (a) research, (b) curriculum theory, and (c) teachers of English and communication. The following is a sample of a few English and reading needs selected from a larger list:

Level I, Grades K-2: The child needs to:

1. Follow simple auditory directions.
2. Develop vocabulary.
3. Discriminate sounds of vowels and consonants.
4. Recite and recognize the alphabet, capital and small letters, printed and cursive.
5. Associate sounds with letter clusters.
6. Write from left to right and top to bottom.
7. Form printed and cursive letters and numbers.
8. Recognize *-s* as plural or person sign.
9. Write declarative and question sentences.
10. Read words and sentences written for Level I with a 70-80% level of comprehension. (This need appears for each Level.)

Level II, Grades 3-4: The child needs to:

1. Demonstrate a 70% performance level on objectives stated for preceding level (Level I).
2. Discuss informally with his classmates.
3. Paraphrase or answer questions on oral reports and presentations.

4. Write sentences, simple paragraphs, and letters.
5. Spell and use words found in reading materials.
6. Recognize dialects of people studied in U.S.A. geography lessons.
7. Retell a story with a logical sequence of events.
8. Read poetry and folk tales.
9. Experience ethnic literature.
10. Write a paraphrase of a pre-read story.

Level III, Grades 5-6: The student needs to:

1. Demonstrate a 70% performance level on objectives stated for preceding level (Level II).
2. Participate and communicate orally in situations such as book reports, discussions, group reports, giving directions, and vocabulary study.
3. Use embedding transformations to produce well formed sentences.
4. Gather and logically organize information.
5. Recount in oral or written discourse folk tales, fables, and myths.
6. Use the dictionary.
7. Use the encyclopedia for information gathering.
8. Use prefixes, suffixes, and root words to expand vocabulary.
9. Recognize a sentence as consisting of a noun phrase and a verb phrase (S — NP + VP) or a subject and a predicate.
10. Write sentences in connected discourse to form paragraphs.

Level IV, Grade 7-8: The student needs to:

1. Demonstrate a 70% performance level on objectives stated for preceding level (Level III).
2. Write and speak nouns, pronouns, verbs, objectives, and adverbs with their proper word forms and markers.
3. Use prepositions and prepositional phrases to add content to sentences.
4. Write subject-verb agreements.
5. Compound and embed sentences to form compound and complex sentences.
6. Spell words using rules of English orthography.
7. Write creatively on topics of interest to students.
8. Speak so that ideas are accurately and logically presented.
9. Select the central idea of written or oral discourse.
10. Separate factual from judgmental statements.

Level V, Grades 9-10: The student needs to:

1. Demonstrate a 70% performance level on objectives stated for preceding level (Level IV).
2. Spell a list of useful words at a 70% performance level.
3. Read materials written at this level at a 70-80% comprehension level.

4. Write a written communication which derives at least a C grade from three experienced and peer graders.
5. Speak an oral communication which derives at least a C grade from three experienced and peer graders.
6. Read three novels from the American, the British, and other ethnic experiences.
7. Identify propoganda, entertainment, and education features of mass media (press, TV, cinema, etc.).
8. Respond to independent reading.
9. Read career oriented discourse.
10. Demonstrate efficiency in basic communications situations such as letter writing, telephone conversation, information gathering, and vocabulary development.

Level VI, Grades 11-12: The students need to:

1. Demonstrate a 70% performance level on objectives stated for preceding level (Level V).
2. Select an English curriculum appropriate to his life goal orientation, e.g., academic, vocational-technical, business, and general.
3. Perform transformations, deletions, and embeddings in generating written sentences.
4. Analyze a written text to determine its content.
5. Demonstrate a 70% comprehension of vocabulary words from a useful vocabulary building list.
6. Identify themes in literature.
7. Respond to works from American, British, and ethnic literature.
8. Demonstrate a knowledge of themes in literature.
9. Write in the forms of the journal and the essay.
10. Participate in large and small group discussions.
11. Read a passage from literature for oral interpretation.
12. Participate as a member of a debate team and/or a forum.
13. Assume the role of a character from fiction or from real life.
14. Record and critically interpret a video tape.
15. Trace themes in literature which parallel history.
16. Demonstrate cultural growth and understanding.

Determining and Developing Objectives—A close look at needs at each level will indicate they contain elements of repetition, excessiveness, and triviality. Not all of the needs will be developed into objectives. Some needs might not be suited to the level; they might not lend themselves to the type of cognitive and skill instruction being discussed here; or they might be trivial. Teachers and consultants will need to thrash out and to sort out a sequence of needs appropriate to each level of instruction, and from this point, develop sets of objectives around concepts or themes.

Objectives in a competency based English curriculum are classified as cognitive and skill objectives, remembering that the competency based portion of the curriculum is, at the most, only about two-thirds of the total curriculum. The affective objectives are the remaining third of the curriculum—supervised, through small group activity, by the teacher. *Cognitive* objectives include (a) concepts, (b) instructional processes, and (c) evaluative processes. *Behavioral* (skill) objectives include (a) behavioral skills and (b) cognitive manipulations. *Affective* objectives include (a) attitudinal changes, (b) opinions expressed, and (c) motivational changes.

Developing Test Items for Objectives—The acid test of a good objective is to be able to write test items for that objective *before* the learning activities are developed. A set of such test items can also be used as a pre-test to determine whether or not students already know the material. If such is the case, then the objective is not useful for that level. For the purpose of individualized instruction, the multiple choice type of questions best suits the purpose. Here is a set of Level V, high school concept and skill objectives and their accompanying test items:

LISTENING—Intonation

Concept objectives:

The students will be able to identify the three features of intonation.

Sample test item:

That quality of speech, intonation, includes all except which of the following features:

a. alliteration
b. stress
c. pitch
d. juncture

LITERATURE—Realism, Expressionism

Skill objective: #1

Describe how the presence of the Stage Manager in "Our Town" is consistent with the character elements of the expressionistic mode of drama.

Sample test item:

The difference in the way the author, T. Wilder, delineates the Stage Manager in "Our Town" from the other characters is:

a. by what they say about themselves to other characters
b. by what the author says about the character
c. by what the character says to the audience
d. by what the other characters say about the character

LANGUAGE

Concept objective:

The students will be able to recognize sentences utilizing commas to set off words in direct address.

Sample test item:

In which sentence is the speaker being derogatory to the person he is addressing?

a. Dope Harry, and George will be upset.
b. Dope, Harry and George will be upset.
c. Dope, Harry, and George will be upset.

SPEAKING—Group Discussion

Concept objective:

The students will be able to arrange the stages of group discussion in solving a problem.

Sample test item:

Of the following steps of group discussion use numbers 1 through 5, one signifying the first step, and 5 the final step, to show the workable pattern to solve a problem.

___ formulate and consider possible solutions
___ state the problem in exact terms
___ analyze data
___ select the best solution
___ collect pertinent data

WRITING—Thematic and Expository Writing

Skill objective #1

The students will be able to order an argumentative paragraph.

Sample test item:

The following sentences compose a complete paragraph. In the space preceding each line mark each line in numerical order (1, 2, 3, 4) with line 1 being the introductory sentence.

___ Many Europeans think we do not.
___ The ordinary American reads cheap literature, mostly from magazines and condensed books—goes not to plays but to movies, usually solely for the sake of "seeing a show" and listens to either "juke box" music of simple strains to folk music.
___ In America, we have a high standard of living and abundant opportunity, but do we have an appreciation of literature?
___ Relatively few Americans read great books, see great plays or listen to great music.

Figure 2 illustrates a strategy for obtaining the needs and objectives for A COMPETENCY BASED ENGLISH CURRICULUM, K-12.

Developing and Evaluating Instructional Materials—Once the objectives and the test items are written, the development and evaluation of these instructional materials will be easy for most teachers. Exemplory instructional materials are available in an abundance of pre-written text materials. No longer will one single "class set" of texts be needed. Some texts might prove inadequate to meet the objectives of this curriculum. Before the finished individualized Learning Activity Packets (LAP) are assembled for inclusion in the completed curriculum at any one level, they should be evaluated by at least three teachers on three different groups of children or students. This type of formative evaluation will assure a usable product.

A Competency Based English Curriculum Guide—A Competency Based English Curriculum Guide will take from two to three years for a dedicated team of teachers in a district to develop. They will need constant assistance and encouragement from their administration and one or more consultants and specialists in English and reading curriculum. The following are the components of such a curriculum guide:

1. There will be one guide for each of the following levels:

Level I,	Grades K-2
Level II,	Grades 3-4
Level III,	Grades 5-6
Level IV,	Grades 7-8
Level V,	Grades 9-10
Level VI,	Grades 11-12

2. There will an Explanatory Statement at the beginning of each guide level which contains the following information:
    a. A definition and rationale for teaching the English, reading, and communication skills.
    b. The diverse and basic needs of students in the ______ School District.
        (1) Vocational needs
        (2) Academic needs
        (3) Social and cultural needs.
    c. Some broad goals for the expected student performance upon leaving the district. What would be a minimal performance level expected of an average, literate, high school educated student from the ______ School District (exit behavior described in broad terms)?
    d. A description of a Competency Based English and Communications Curriculum.
        (1) It provides for individual language abilities and development for students.
        (2) It provides for individual vocational, academic, and cultural needs and interests of students.

(3) It assures that specific skills and knowledge are obtained by each individual as he progresses from level to level.
(4) It provides individualized learning activities for each student.
(5) It provides an enter and an exit test at each level to assure the teachers and students that at least 70% mastery level at each level.
(6) It provides for diagnosis so that students can review skills and knowledge of materials at a lower level if mastery has not been at a satisfactory level.
(7) It provides times and facilities for teacher and student to work creatively in an open-classroom setting on useful and thoughtful communications tasks in the area of language, literature, and composition involving listening, speaking, reading, and writing skills.

3. The body of each guide at a level will be unique and will contain the following:
   a. *A statement of specific English, reading, and communications needs* of students at the grade and age level. These needs should reflect the average language development of children and students at the particular level.
   b. *A statement of broad objectives for each level.* These should be selective and represent skills as well as concepts. They should cover language, literature, and composition and include communication skills of listening, speaking, reading, writing, and media.
   c. *A sample entrance and exit test for each level.*
      (1) Both the entrance and exit test should be the same except for slight changes in the items. They are a form A and B of the same test.
      (2) The items are obtained from the specific objective.
   d. *The sample objective, activity, test item for each level.* This section should grow and change from year to year.
   e. *Samples of individualized Learning Activity Packets for each level.* This section should grow and change from year to year.
   f. *Suggested outline for mini-courses, open-classroom activities, and creative learning activities.* This section would include such items as activities for groups, learning centers, units, and mini-courses and can suggest audio-visual, literature, and writing activities. This section should grow and change from year to year.
   g. *Bibliography of teacher references, student textbooks, literature, and audio-visual materials.* This section should grow and change from year to year.

Figure 1

DEVELOPMENTAL STAGES OF A COMPETENCY BASED ENGLISH AND COMMUNICATIONS CURRICULUM, GRADES K-12

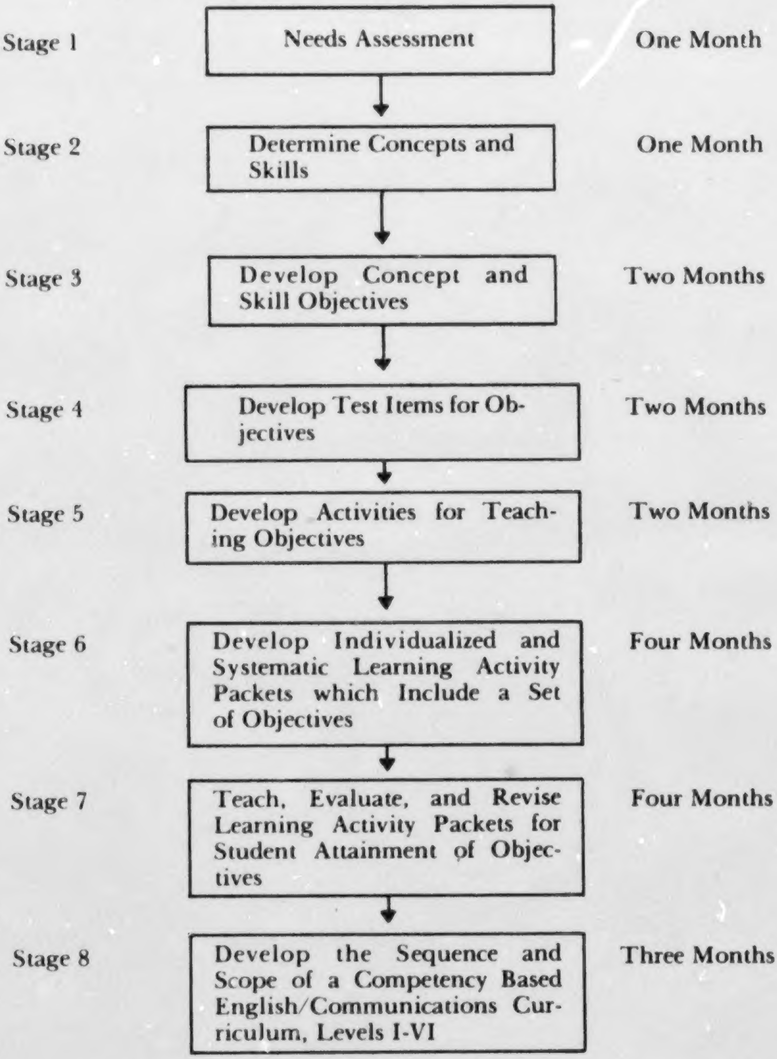
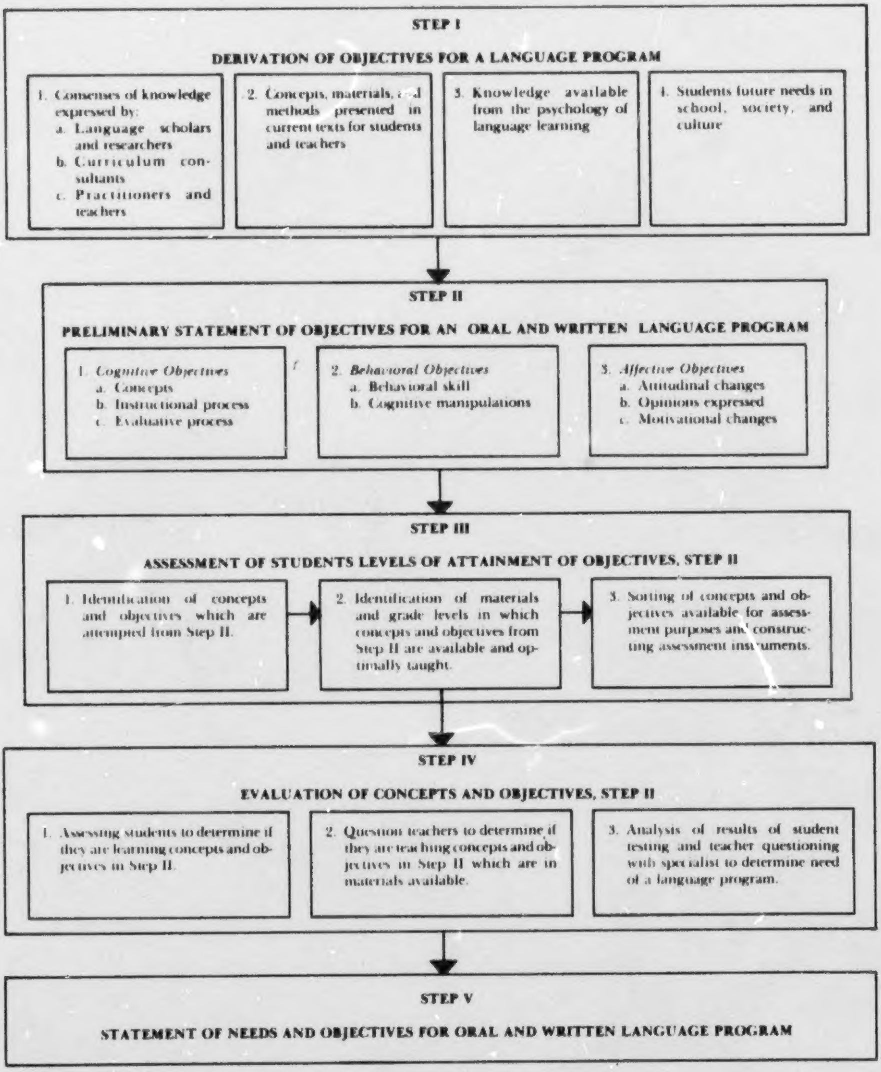

Stage 1
Needs Assessment
One Month
Stage 2
Determine Concepts and Skills
One Month
Stage 3
Develop Concept and Skill Objectives
Two Months
Stage 4
Develop Test Items for Objectives
Two Months
Stage 5
Develop Activities for Teaching Objectives
Two Months
Stage 6
Develop Individualized and Systematic Learning Activity Packets which Include a Set of Objectives
Four Months
Stage 7
Teach, Evaluate, and Revise Learning Activity Packets for Student Attainment of Objectives
Four Months
Stage 8
Develop the Sequence and Scope of a Competency Based English/Communications Curriculum, Levels I-VI
Three Months

Figure 2

A Strategy for Obtaining the Needs and Objectives for a Competency Based English Curriculum, Grades K-12

Stages 1, 2, 3, and 4



STEP I

DERIVATION OF OBJECTIVES FOR A LANGUAGE PROGRAM

1. Consenses of knowledge expressed by:
   a. Language scholars and researchers
   b. Curriculum consultants
   c. Practitioners and teachers

2. Concepts, materials, and methods presented in current texts for students and teachers

3. Knowledge available from the psychology of language learning

4. Students future needs in school, society, and culture

STEP II

PRELIMINARY STATEMENT OF OBJECTIVES FOR AN ORAL AND WRITTEN LANGUAGE PROGRAM

1. *Cognitive Objectives*
   a. Concepts
   b. Instructional process
   c. Evaluative process

2. *Behavioral Objectives*
   a. Behavioral skill
   b. Cognitive manipulations

3. *Affective Objectives*
   a. Attitudinal changes
   b. Opinions expressed
   c. Motivational changes

STEP III

ASSESSMENT OF STUDENTS LEVELS OF ATTAINMENT OF OBJECTIVES, STEP II

1. Identification of concepts and objectives which are attempted from Step II.

2. Identification of materials and grade levels in which concepts and objectives from Step II are available and optimally taught.

3. Sorting of concepts and objectives available for assessment purposes and constructing assessment instruments.

STEP IV

EVALUATION OF CONCEPTS AND OBJECTIVES, STEP II

1. Assessing students to determine if they are learning concepts and objectives in Step II.

2. Question teachers to determine if they are teaching concepts and objectives in Step II which are in materials available.

3. Analysis of results of student testing and teacher questioning with specialist to determine need of a language program.

STEP V

STATEMENT OF NEEDS AND OBJECTIVES FOR ORAL AND WRITTEN LANGUAGE PROGRAM

The Pennsylvania State University

Improvisation and the Teaching of English

by

Patrick L. Courts

When I finished preparing my talk for today, I realized that I had violated just about every rule we lay out for the production of discourse when we are telling students how to produce speeches or papers or whatever. For example, I didn't begin with an outline; I didn't define my topic or write down a statement of theme; I even forgot to bring a pen or sheet of paper to the place I was sitting. On top of all these things I didn't do, sins of omission, so to speak, I also did some bad things because I was sitting in a very comfortable chair and listening to my collection of Bob Dylan records. At the same time, I was thinking over the conferences I've been to this past year, the talks I've heard, the teachers I've talked to, and the students I've been teaching.

While I was involved in this somewhat muddled thinking process, I began to fall in and out of a daydream I have every once in a while, wherein I am at a party sometime in the future, and the theme of the party is the 1960's. People are dressed in mixtures of army fatigues from army surplus stores and old, loose fitting clothing from salvation army stores. The prize for the best costume is the soundtrack from the movie *Woodstock*, and other prizes are being given for the most authentic protest signs. Most of the people are milling around "doing their own thing," though some people are made up in blackface and pretending they are "sitting in" over near the kitchen. Everyone is having a good time listening to early Beatles records, Dylan, and the Jefferson Airplane; and the conversations consist mainly of people reminiscing about the Civil Rights Movement, the 1968 Democratic Convention, and how Mayor Daley was the last of the great, old politicians. The conversations also abound with phrases like "ripping stuff off," "offing pigs," and "blowing people's minds." Everyone, however, is careful to avoid serious discussions of things like the Watts riots, the invasion of Cambodia, the ensuing deaths at Kent State, and the assassinations of the Kennedys and Martin Luther King. After all, it is a party.

Anyway, in the midst of this somewhat nightmarish vision and, at the same time, trying to think of exactly what I was going to talk about today, two lines from Bob Dylan's far-ranging repertoire kept imposing themselves on my now thoroughly overwrought consciousness. (I was beginning to think that maybe I should have begun with an outline.) The first line that popped into my head is from a Dylan song called "The Ballad of a Thin Man." It is a song about a man who is trapped in a world he no longer understands. He sees the ferment around him, the confusion, and the arguments, but he cannot or will not understand. It almost seems as though he has been taught to ignore the realities around him, taught that he cannot really affect these realities anyway, and that he has learned his lesson well. The line I'm thinking of forms a kind of chorus in the song, and it is, "Something is happening, here, but you don't know what it is, do you, Mr. Jones?" In short, Mr. Jones would quite likely be at a party like the one I just described.

The other line that popped into my head comes from a song titled "Subterranean Homesick Blues"; it says quite matter of factly that "You don't need a weatherman to know which way the wind blows." Some of you may know that this is the line from which the Weathermen, a radical political group of the 60's, chose their name. But the line is much more important than this bit of trivia suggests because it implies that although things sometimes seem very complex and confused, so much so in fact that we are rendered impotent and unable to filter out a course of personal action, that even amidst this confusion, things are often a lot clearer than they sometimes seem. It really isn't very difficult to figure out which way the wind is blowing.

Of course, when Dylan wrote these lines he did not have education in general or English teaching in particular on his mind, but the lines do apply nicely to us and our profession. We live in a time of ferment and controversy, and as professional educators we are bombarded with jargon-laden arguments for and against behavioral objectives, accountability, performance-based certification, specification of teacher competencies, improved interface, humanistic curricula, flexible instructional systems, standardized tests for language arts, *ad nauseam*. Here in New York, the situation is being further confused by the appointment of an Inspector General of Education whose job it is to make education in this state more efficient, probably by implementing systems approaches to the teaching of English based on the behavioral systems used by the Pentagon in the sixties to design planes that would not fly and later discarded by the Pentagon because of the excessive cost and waste incurred by the system.

Clearly, something is happening here, and it would take someone with Mr. Jones's blinders to avoid seeing what it is. We do not need a weatherman to know that most of the wind that's been blowing in education lately has been concerned, not with what we actually do in our classrooms, not with what a student should learn in an English class, not with the activities that make up this thing we call English; but with how to measure what we do, how to account for it, how to make it less expensive. In short, we have been looking at things backwards by avoiding the very difficult questions of what and how to teach, by making the very dangerous assumption that learning is measurable and quantifiable, and by operating on that assumption in face of a growing body of information that strongly suggests that almost all IQ tests, reading tests, and writing tests are severely limited and discriminatory in terms of race, culture, class, dialect, and geographic locale.

We don't need a weatherman to know that the teaching of grammar (when that means to study *about* language rather than to work with the varieties of language we and our students produce) does not change a person's competence in that language: he will not talk "better," read "better," or write "better" as a result of the twelve years of grammar he studies in elementary and high school any more than that freshman composition course based on grammar and usage will change his life. And we don't need a weatherman to know that many of our students are unable to read and even more don't much like to read because we have cut them off from the topics and books that have immediate bearing on their lives. Instead, we have rigidly enforced a traditional curriculum emphasizing the transmission of our cultural heritage through the study of great books and the writing of literary criticism about these great books. At its best, such a system produces literary critics who will go on and do the same thing to their students; at its worst, it produces people who hate literature and who think that they must have an English teacher present before they could ever decipher the "hidden" meanings of a poem.

But I have been negative enough, and it is only fair to say that you don't need a weatherman to know that the winds of change have blown some fresh air into the teaching of English. Furthermore, it is about time that I heeded my own call and started talking, specifically, about some of the concrete things I think teachers should be doing in their English classrooms.

And so, I finally come to the topic that was listed for my presentation, Improvisation and the Teaching of English. I chose this topic because I believe that improvisation can profitably serve as one of the central activities of the English class—an activity of major importance to the teaching of writing, literature, talking, and human experience in general.

I suppose the first order of business, then, is for me to define what I mean by improvisation and to offer some concrete examples of how improvisations can operate in English classrooms. Improvisation can involve students in the relatively simple but fun activity of pantomiming words like *anger, love, pride, fear*, etc., while the rest of the class or group tries to guess what it is the actor is trying to convey. Or, the pantomime can be made slightly more complex by asking students to act within the context of common situations: a man has lost his last coin in a cigarette machine and has still not gotten his cigarettes; a person taking a drink from a water fountain gets drenched because the water pressure is too high; a fisherman loses the "big" one. The situations themselves are fun and easy to create, and I've yet to find students who would not participate happily in these kinds of activities. Aside from the fact that students generally enjoy doing these pantomimes and trying to guess what word or situation is being pantomimed, this activity also provides a concrete base for a variety of writing activities. At the most basic level, students can write short pieces describing the situation that was pantomimed, or they might wish to write about a time they had trouble with a Coke machine or the time they lost the "big" one. They might wish to write a criticism of the person who performed the pantomime, suggesting actions he should have employed to better convey his word or situation. Other students may wish to go so far as to write a short story about the situation they saw pantomimed, or a character sketch of a person they know who is *angry, in love, prideful,* etc. In my own mind, the most important thing here is not the kind of composition they write, but that they have a concrete experience to base their writing on.

At a slightly more complex level, students can improvise situations of their own devising in which several people take part and in which they may or may not use words, depending on their own wishes. For example, they might do a take-off on the man who lost his money in the cigarette machine. Just when he is at his angriest and kicking and shaking the machine the hardest, the store owner or hotel clerk might approach him and politely (or angrily) request that he immediately stop wrecking the machine. The possibilities of such a scene should be obvious: as the two men begin to argue hotly, a policeman comes along, the first man's wife or child might come along, and for that matter, a Martian might come along. All that really matters is that the people continue to interact in terms of the initial conflict, possibly taking sides for and against the machine.

Again, such an activity suggests a wide variety of writing activities: some students might be enterprising enough to write a one-act play based on the improvisation, a play they might later perform and possibly even videotape. Students involved in this kind of follow-up activity could also work out costuming, background music, and even an integrated slide presentation, if they wished to pursue the idea to its natural end. Other students might only wish to write character sketches of the people in the improvisation, or expository papers describing exactly what happened in the improvisation, possibly working from the journalist's objective point of view. And still others might write persuasive essays taking the part of one side or another (for or against the machine) and trying to convince the reader of their point of view. Ideally, of course, students will create their own conflict situations, and with a few simple directions from the teacher, they will have little difficulty in creating, planning, and when they wish, performing their improvisations.

In preparing students for these kinds of improvisations, the teacher can employ a variety of strategies, but I prefer to break the class into groups of four or five students and to ask them to think about the kinds of conflicts they see everyday or in which they, themselves, are sometimes involved. A few examples will help them here: students might improvise a situation in which a son or daughter brings home the "undesirable" person he or she has been dating against his parents wishes. Such a situation might resolve itself when the parents find out that the undesirable is a nice person after all, or when everyone becomes so angry that the two young people leave the house. Other conflict situations that are relatively familiar to many of our students are those that deal with racial, ethnic, or religious bias, those that deal with the generation gap, and those that focus on controversial topics like women's liberation, ecology, student activism, etc.

At any rate, I've usually found that as long as the students have a clear idea of the kinds of conflicts they might improvise, they have little difficulty creating their own. The only direction I generally give them when they are planning these kinds of improvisations is that they should carefully plan an ending for their improvisation. This will help them to avoid the deadly improvisations which seem to drone on forever, doing the same things over and over again, because no one involved in the improvisation knows how to get out of it. Another important generalization to consider is that the performance of these improvisations might be the least important and necessary part. It is the creating and planning process that should be emphasized because it is there that the students are actively engaged in creating literature and struggling with the problems of organization, characterization, stereotyping, plot, theme, etc.

In my own experiments with improvisation, I have found that most students enjoy the planning stage, but some of them become terrified by the acting-out stage. At the same time, there are always a few who are happy and willing to perform their improvisations, and these few groups supply us with all we need for large group discussions. And as far as writing activities are concerned, students who have not performed their improvisations can still write about the ones they see other students perform; they can write about the improvisation they planned in their group; or, they might want to explain why it is they are afraid to perform their own improvisation. Some may want to script their improvisation and see if another group will perform it. The point is that we don't have to torture anyone with improvisations.

Earlier in this discussion, I said that I viewed improvisation as the central activity of the English class, out of which a variety of other activities should grow. So far, it may seem that I have said a lot about improvisation and writing, but very little, if anything, about literature. Quite the contrary, though. I think I have said a great deal about literature for the improvisations themselves are a dramatic form of literature and no less valuable because they have been created by students. On the other hand, we all are interested in teaching professional literature, and I think that improvisations can provide a lot of help here also.

If, when we teach literature, it is our desire to help our students discover their own and others' humanness, and if we see literature as the man's artistic expression of the human condition, of the consciousness and the subconscious, as the artistic expression of those private moments we all have but never articulate, as the catalogue of events which make us a part of that group we call humanity—if we see literature in these terms, then improvisation has a natural relationship to the study of literature. The themes, the conflicts, the funny little incidents, and the human interactions which are at the center of dramatic improvisations are also at the center of literature. And what better way to begin to experience the humanness of literature than to participate in some concrete representations of the themes, emotions, and conflicts presented by the writer.

Before I go any further with this point, I should be clearer about the kinds of improvisations I see students doing in preparation for or response to literary works. It is a common mistake to ask students to perform a scene taken straight out of a novel or dramatic poem, and the reason it is a mistake is that it limits the students too much. Instead of exploring the nature of the emotions, the conflicts, and/or the interactions with which the work deals, they are put in the unpleasant position of copying or re-producing what has already been done. I give this caution only because I've seen so many students and teachers suffer through these moments, all in the best of faith. I suggest that it is more worthwhile for the students to create their own improvisations based on the *kinds* of conflicts represented in a novel, or on the kinds of human relationships, or kinds of situations. For example, if the novel is *Grapes of Wrath*, students might work out improvisations dealing with poverty, a family breaking up, or the hardships of farm life or of being migrant workers. This kind of activity is valuable because it gives the students an opportunity to explore their own ideas and feelings about the themes and topics presented in the work of literature, and the improvisational experience offers a concrete base for comparisons between the way the students see things and the way the writer presents things. The literature has bearing on the lives and experiences of the students because the classroom activities and approaches to the literature are focused on the students and their ideas.

Improvisations offer the students a chance to bring the real events of their lives into the classroom for discussion, analysis, comparison, and to serve as beginning points for writing and the study of literature. An emphasis on improvisation in the classroom puts English back into a living context and removes it from the sterile, dusty place it so often holds in the schools. But even more important, improvisations, by their very nature, stress the interests of the students and they stress action. The students create, plan, and sometimes perform improvisations. They produce, direct, and make costumes and background music for videotape productions of their improvisations. They write about their improvisations, talk about them, argue and laugh about them. They read about the themes and conflicts which have formed the central conflicts in their improvisations. In short, because of improvisations, they *do* things with language: instead of only studying about English, they language about it. They use English to understand themselves, their world, and their place in it. They use English to share themselves with their classmates, and, through the reading they take part, emotionally and intellectually, in man's greatness, his pettiness, his sorrow and his joy.

I think that there is little question that such an approach is different from the traditional approach, and that, while it may not produce what we generally consider literary critics, it probably will generate students who have some understanding of their own worth as human beings, who have self-confidence in their ability to speak and write, and who enjoy reading.

State University College
Fredonia

Literary Research

by
R. D. Lakin

I often think of Emily,
who knew the true uses of sound
and never went gathering chatter
on an endless round of calls,
a matter for modern comment.
I myself can see her baking pies
or brightening the wash
accompanied by buzzing flies,
but a woman's avoiding hairdressers
suggests a quirk or even worse.
(Only confessors dare dogmatize.)
My theory is a fabulous Turk,
unknown so far to American Lit,
crept in at night and undid
her hair in tresses of dark beauty
she stroked in lonely pride,
then bid him do his duty
and pin it back in a bun so staid
her own sister made no note
she'd ever been undone.

Fort Collins, Colorado

MEDIA REVIEW

"The Life and Times of Judge Roy Bean." Distributed by Swank Motion Pictures, Inc., 201 Jefferson Ave., St. Louis, Missouri, 63166. Running time: 120 minutes. Rental: $125.00. (PG). Color.

"A man has two loves in his life—an unattainable goddess and a mortal woman." So states Judge Roy Bean in this rich mélange of fact, fiction, love story, spoof, history, satire. Bean, a man like Roosevelt suited to the times and the land, appears in many roles: he is cooly outrageous as the Judge of the town of Vinegaroon ("Sting of a Tarantula's Tail") and casually eliminates opposition between hands of a card game, assured of the rectitude of his actions because of his nodding acquaintance with Psalm 58, "The righteous will rejoice when he sees the vengeance." Yet, like some raw Don Quixote of the raucous West, he can follow in quest of a dream of the ideal woman, Miss Lily Langtry. With gentle, stilted gallantry, he writes to her a moment before death, "Your presence on the earth has given me strength and dignity to become unto a gentleman. God willing I might sometime stand in your light." He appears as the romantic sensible to the "positively tropical" beauty of his land, and as the aging man who sees his dream quietly slipping away and fears, "People won't know what it took to build this place; they won't know about me. . . Man is mortal." Finally, he is the rough-cast Odysseus who returns to rid the town of the pervasive decay that spread during his absence; like a mythic hero out of some old picture book, fighting a rearing horse and crying, "For Texas and Miss Lily!", he achieves his final definition as "Justice" in the apocalyptic explosion that purifies the town.

The action and dialogue excited both non-Regents and Regents students, and the film lent itself to a variety of thematic approaches: the hero as rebel; love as real and ideal; the American Dream and its intimate indentification with the land; the opposition of vengeance and justice. It is an obvious resource for courses in The Cowboy and American Humor, containing such folk types as the original Bad Bob, the Albino, and Grizzly Adams, who "cohabited with the bears." Highly recommended for Senior High School. Publicity and free bulletin board posters available from Swank.

L. Edison

"Overture/Nyitany." Running time: 9 minutes. Rental: $12.50. Distributed by Contemporary McGraw Hill, Princeton Road, Hightown, New York 08520. Color.

"Overture/Nyitany" is a superb film for creative writing. Using special techniques, it reveals the development of an embryo from the first cellular stirrings to the final emergence of a chicken from an egg. The processes of growth are recorded against a splendid background of sound and color, and the classical music provides an amazingly accurate accompaniment to the events.

The film is a rich resource for poetry and descriptive writing assignments. It provides an excellent visual example of the idea of the microcosm and can be used to explain the different types of narrators in fiction. It is also highly suitable for courses in Science Fiction, especially in connection with the theme of fantastic voyages and the archetypal pattern of the journey.

Highly recommended for all levels.

L. Edison

"Between Time and Timbuktu." Running time: 9 minutes. Rental: $75.00. Distributed by New Line Cinema, Educational Film Division, 121 University Place, New York, New York 10003.

Strands of "Cat's Cradle," "Welcome to the Monkey House," "Sirens of Titan," "Happy Birthday Wanda June" and "Slaughter-House Five" are woven into this Vonnegut composite. The film's protagonist is Stony Stevenson, resident of Indianapolis and winner of the Blast-Off Space Food Jingle Contest; his journey through the Chrono-Synclastic Infundibulum and, at various and unannounced times, back to earth, provides the framework for the glancing Vonnegut view of contemporary society. Stony's quest for the meaning of life and death climaxes in an Ethical Suicide Parlor where a customer who has "dropped in to turn off" asks, "What are people for, anyway?" Not surprisingly, there is no answer to this old man's last question.

The film is useful for classes in satire and science fiction, and for units on mythology, archetypal patterns, contemporary writers and values. Recommended for High School.

L. Edison

"Dream of Wild Horses." Running time: 9 minutes. Rental: $12.50. Distributed by Contemporary McGraw Hill, Princeton Road, New York 08520.

This remarkable film takes as its subject the wild horses of Camarque. The use of slow motion and soft-focus backgrounds turn the events of the film into scenes of haunting beauty. The audience is absorbed by the life of the herd—stallions brutally fight for dominance, the herd gallops through ocean mist in the early dawn, the animals race through fire. At times, the film makes us see the herd as nearly a pure abstraction of rhythm and movement.

This film, an Award of Merit winner at the 1960 Edinburg Film Festival, is a rich source for students of creative writing of all tracts—special, non-regents, regents, advanced. It also lends itself well to units on the West. Highly recommended for all levels.

L. Edison

"Blaze Glory." Running time: 12 minutes. Rental: $15.00. Distributed by the National Education Film Center, Route 2, Finksburg, Maryland 21048. Color.

This spoof on the weary TV western is both instructive and highly entertaining. The plot, over-packed with the typical elements of the western, moves quickly through a series of events which provide, with the help of special effects, an animated definition of parody. The film's presentation of the "cowboy-hero" in the ridiculous character of Blaze Glory quickly elicits discussion of our expectations of the hero in the traditional western flick.

In addition to its obvious use in a unit on sature, it could also well be shown in units on the hero, the literature of the American West, and, of course, in humor.

J. Feighery.